Volume 20, No. 7

July 1959

АВТОМАТИКА
и
ТЕЛЕМЕХАНИКА

SOVIET INSTRUMENTATION AND
CONTROL TRANSLATION SERIES

Automation and Remote Control

(The Soviet Journal *Avtomatika i Telemekhanika* in English Translation)

■ This translation of a Soviet journal on automatic control is published as a service to American science and industry. It is sponsored by the Instrument Society of America under a grant in aid from the National Science Foundation, continuing a program initiated by the Massachusetts Institute of Technology.

SOVIET INSTRUMENTATION AND CONTROL TRANSLATION SERIES

■ This translation of the Soviet Journal *Avtomatika i Telemekhanika* is published and distributed at nominal subscription rates under a grant in aid to the Instrument Society of America from the National Science Foundation. This translated journal, and others in the Series (see back cover), will enable American scientists and engineers to be informed of work in the fields of instrumentation, measurement techniques, and automatic control reported in the Soviet Union.

The original Russian articles are translated by competent technical personnel. The translations are on a cover-to-cover basis, permitting readers to appraise for themselves the scope, status,and importance of the Soviet work.

Publication of *Avtomatika i Telemekhanika* in English translation started under the present auspices in April, 1958, with Russian Vol. 18, No. 1 of January, 1957. Translations of Vols. 18 and 19 have now been completed. Translation of Vol. 20, No. 1 will be published in January 1960, and the twelve issues of Vol. 20 will be published in English translation by October, 1960.

Transliteration of the names of Russian authors follows the system known as the British Standard. This system has recently achieved wide adoption in the United Kingdom, and is being adopted in 1959 by a large number of scientific journals in the United States.

All views expressed in the translated material are intended to be those of the original authors, and not those of the translators, nor the Instrument Society of America.

Readers are invited to submit communications on the quality of the translations and the content of the articles to ISA headquarters. Pertinent correspondence will be published in the "Letters" section of the ISA Journal. Space will also be made available in the ISA Journal for such replies as may be received from Russian authors to comments or questions by American readers.

1959 Volume 20 Subscription Prices:

Per year (12 issues), starting with Vol. 20, No. 1

General:	*United States and Canada*	$35.00
	Elsewhere	38.00
Libraries of nonprofit academic institutions:		
	United States and Canada	$17.50
	Elsewhere	20.50
Single issues to everyone, each		$ 6.00

1957 Volume 18 and 1958 Volume 19 issues also available. Prices upon request.

See back cover for combined subscription to entire Series.

Subscriptions and requests for information on back issues should be addressed to the:

Instrument Society of America
313 Sixth Avenue, Pittsburgh 22, Penna.

Translated and printed by Consultants Bureau, Inc.

Volume XX No. 7 July 1959

English Translation Published March 1960

Automation and Remote Control

The Soviet Journal Avtomatika i Telemekhanika in English Translation

Reported circulation of the Russian original 8,000.

Avtomatika i Telemekhanika is a Publication of the Academy of Sciences of the USSR

See following page for Table of Contents.

AUTOMATION AND REMOTE CONTROL

Volume 20, Number 7 July 1959

CONTENTS

CONTENTS (continued)

A GENERAL CONDITION FOR AN EXTREMUM OF A GIVEN FUNCTION OF THE MEAN-SQUARE ERROR AND THE SQUARED MATHEMATICAL EXPECTATION OF THE ERROR OF A DYNAMIC SYSTEM

N. I. Andreev

(Khar'kov)

The derivation is given for the condition for an extremum, and also for the greatest or least value, of some function f of the mean-square error and the squared mathematical expectation of error in approximating a random function. The general condition obtained is applied to the problem of choosing an optimal nonlinear integral operator. The paper is based on the ideas and results presented in [1-4].

The determination of the optimal characteristics of an automatic control system is usually connected with the necessity of finding an optimal system operator which provides an extremum of some criterion. In many practical problems it is advantageous to choose as the criterion for comparing various systems some given function, $f_1 = f_1(p, D)$, of the squared mathematical expectation, p, and the dispersion, D, of the error at the system's output.

In what follows it is more convenient to work, not with the dispersion, but with the mean-square error, which we shall denote by q. The function $f_1(p,D)$ is transformed correspondingly to the function $f(p,q)$.

It is interesting to pose the problem in the following manner. As the result of observations, let it be required to approximate the random function X(t) by the random function Y(t). In order to implement the approximating function Y(t) as a result of observations on X(t), it is necessary that the random function X(t) be subjected to some transformation (for example, be transformed by means of some automatic control system). If we denote the operator of the transformation by A, we can give the following formulation of the problem of approximating the random function: to find such an operator A for which F)p,q) will attain an extremum

$$f(p, q) = f_{\text{extr}}, \tag{1}$$

where

$$p = \{M[Y(s) - AX(t)]\}^2, \quad q = M\{[Y(s) - AX(t)]^2\}. \tag{2}$$

Relationship (1) must hold for all values of the variable s lying in some region S.

For each given value of the argument s, the random functions Y(s) and AX(t) are random variables. It is therefore advantageous to first consider the simpler problem of approximating the random variable Y by the random variable Z lying in a given set, L, of random variables which does not contain Y. For this, we assume that the set L is a linear space, i.e., if $Z_1, \ldots, Z_n$ are random variables lying in the set L, then the random

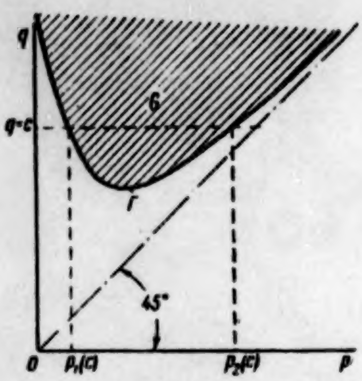


Fig. 1.

variable $Z = c_1Z_1 + \ldots + c_nZ_n$ lies in the set L for any $\underline{n}$ and for any nonrandom numbers $c_1,\ldots,c_n$. We then pose the problem: to find that random variable $Z_n \in L$*, for which

$$f(p, q) = f_{\text{extr}},$$

where

$$p = [M(Y - Z)]^2, \; q = M[(Y - Z)^2].$$

The values of the variables $\underline{p}$ and $\underline{q}$ vary within some region G which, for a given random variable Y, is defined by characteristics of the set L of random variables Z.

The function $\underline{f}$ may have extrema inside region G, and greatest and least values on the boundary of G. The values p_e and q_e for which function $\underline{f}$ attains extrema are determined as solutions of the system of equations

$$\frac{\partial f}{\partial p} = 0 \quad \frac{\partial f}{\partial q} = 0, \tag{3}$$

if the function $\underline{f}$ is differentiable with respect to $\underline{p}$ and $\underline{q}$. Conditions (3) are necessary conditions that the function $\underline{f}$ attain an extremum for $p = p_e$ and $q = q_e$. The function $\underline{f}$ can also reach extrema at points inside region G at which one, or both, of the partial derivatives $\partial f/\partial p$, $\partial f\ \partial q$/does not exist.

In seeking a solution of the system of equations in (3), we may write the necessary conditions for an extremum of the function $\underline{f}$ in the form

$$p = p_e, \; q = q_e, \tag{3a}$$

where p_e and q_e are known constants.

In the general case, equations (3a) do not have unique solutions.

In practical problems, great interest usually inheres in the greatest, or least, values of the function $\underline{f}$. To find these values of the function $\underline{f}$, we first determine the boundary, Γ, of region C (see Fig. 1). For this, we take several values of the mean-square error $q = c$ and for these fixed values of $\underline{q}$ we determine the extremum (extrema) of the squared mathematical expectation $p = p_i(c)$, $i = 1,\ldots,k$ (Cf. Fig. 1). The values $q = c$, $p = p_i(c)$ have corresponding to them the points $\Gamma(c)$ on the boundary Γ. By giving different values to the parameter $\underline{c}$, one can obtain all the points on the boundary Γ. The problem of finding the extremum $\underline{p}$ with the condition $q = c$ reduces, in our case [5], to the problem of finding the unconditional extremum of the sum

$$p + \lambda q, \tag{4}$$

where λ is defined by the condition $q(\lambda) = c$.

In the problem here, it is not necessary to seek the values of the artificially introduced parameter λ, corresponding to $q = c$, since we are not interested in any given concrete value of the dispersion $\underline{q}$. It is necessary for us to find the relationship giving $\underline{p}$ and $\underline{q}$ as functions of λ along the boundary Γ of region G. This relationship will be found if we obtain a method for finding the extrema of expression (4) for various values of the parameter λ. After the relationships $p = p(\lambda)$ and $q = q(\lambda)$ have been found, the greatest and least values of the function $\underline{f}$ are determined as extrema of the function $\Phi(\lambda)$:

$$\Phi(\lambda) = f[p(\lambda), q(\lambda)]. \tag{5}$$

* The notation $Z \in L$ means that Z lies in L (Z belongs to the set L).

Seeking the extrema of the function Φ is done by the ordinary methods of mathematical analysis.

Thus, in order to determine the boundary Γ of the region G, it is necessary to develop a method of determining extrema of the sum $p + \lambda q$, where λ is a parameter whose region of variation depends on the conditions of the concrete problem.

We first consider the case $\lambda \geq 0$, and obtain the corresponding condition for the minimum of the sum $p + \lambda q$. For this, it is necessary to find that random variable $Z \in L$ for which the inequality

$$[M(Y-Z)]^2 + \lambda M[(Y-Z)^2] \leqslant [M(Y-U)]^2 + \lambda M[(Y-U)^2] \tag{6}$$

holds for any U L.

By following the method of V. S. Pugachev [1, 2], we consider the equation

$$\begin{aligned}[M(Y-U)]^2 + \lambda M[(Y-U)^2] = [M(Y-Z)]^2 + \\ + \lambda M[(Y-Z)^2] + [M(Z-U)]^2 + \lambda M[(Z-U)^2] + \\ + 2M(Y-Z)M(Z-U) + 2\lambda M[(Y-Z)(Z-U)],\end{aligned} \tag{7}$$

which is valid for any random variables Y, Z and U. Let there be a random variable Z such that, for all random variables $U \in L$,

$$M(Y-Z)M(Z-U) + \lambda M[(Y-Z)(Z-U)] = 0. \tag{8}$$

Then, for any random variable $U \in L$, equation (7) takes the form

$$\begin{aligned}[M(Y-U)]^2 + \lambda M[(Y-U)^2] = [M(Y-Z)]^2 + \\ + \lambda M[(Y-Z)^2] + [M(Z-U)]^2 + \lambda M[(Z-U)^2]\end{aligned} \tag{9}$$

For $\lambda \geq 0$, inequality (6) follows from this equation since, in this case,

$$[M(Z-U)]^2 + \lambda M[(Z-U)^2] \geqslant 0.$$

Thus, for $\lambda \geq 0$, expression (8) is a sufficient condition that inequality (6) hold.

We now prove that condition (8) is also necessary for inequality (6) to hold for all $U \in L$. We assume that, for some random variable $U_1 \in L$,

$$M(Y-Z)M(Z-U_1) + \lambda M[(Y-Z)(Z-U_1)] \neq 0. \tag{10}$$

We set

$$U_\alpha = Z + \alpha(U_1 - Z). \tag{11}$$

where α is an arbitrary parameter. The random variable U_α lies in space L since, by definition, L is a linear space.

For U_α we have

$$\begin{aligned}M(Z-U_\alpha) = \alpha M(Z-U_1), \quad M[(Z-U_\alpha)^2] = \alpha^2 M[(Z-U_1)^2]. \\ M(Y-Z)M(Z-U_\alpha) + \lambda M[(Y-Z)(Z-U_\alpha)] = \\ = \alpha\{M(Y-Z)M(Z-U_1) + \lambda M[(Y-Z)(Z-U_1)]\}.\end{aligned}$$

Equation (7) takes the form:

$$[M(Y-U_\alpha)]^2+\lambda M[(Y-U_\alpha)^2]=[M(Y-Z)]^2+\lambda M[(Y-Z)^2]+$$
$$+\alpha^2\{[M(Z-U_1)]^2+\lambda M[(Z-U_1)^2]\}+$$
$$+2\alpha\{M(Y-Z)M(Z-U_1)+\lambda M[(Y-Z)(Z-U_1)]\}. \tag{12}$$

It is clear from this that, for all values of the parameter α which are opposite in sign to the quantity

$$M(Y-Z)M(Z-U_1)+\lambda M[(Y-Z)(Z-U_1)]$$

and which are sufficiently small,

$$[M(Y-U_\alpha)]^2+\lambda M[(Y-U_\alpha)^2]<[M(Y-Z)]^2+\lambda M[(Y-Z)^2].$$

This proves that condition (8) is necessary for inequality (6) to hold for any $U \in L$.

We have thus proved that, for $\lambda \geq 0$, condition (8) is a necessary and sufficient condition for a minimum of the sum $p + \lambda q$ when the random variable Y is approximated by a random variable $Z \in L$.

Equation (8) may be rewritten in the form

$$M(Y-Z)MV+\lambda M[(Y-Z)V]=0. \tag{13}$$

where V is an arbitrary random variable of the set L.

It is easily seen [1, 2] that if the set L contains two random variables Z_1 and Z_2 which satisfy condition (8) for all random variables $U_\alpha \in L$ then, with probability one, these random variables are equal to each other.

We now consider the random function X(t) and some set R of operators which transform functions of the argument $\underline{t}$ into functions of the argument $\underline{s}$. The arguments $\underline{t}$ and $\underline{s}$ vary, respectively, in the region T and S. As a result of transforming the random function X(t) by all the operators of the set R one obtains a set of random functions of the argument $\underline{s}$ and, for each given value of the argument $\underline{s}$, a set of random variables. In order that the results obtained above be applicable to this set of random variables, it is necessary that this set be a linear space. For this, the set of operators must itself be a linear space (Cf. [2]).

We now return to the problem posed at the beginning of this paper: to find, in a given class of operators, an operator A such that condition (1) will hold for each given value of $\underline{s}$, where it is assumed that the set of operators, lying in a given class R, forms a linear space. In the case, of great practical significance, when the function $\underline{f}$ attains a greatest, or least, value on the boundary of the region of variation of $\underline{p}$ and $\underline{q}$, the first step in the search for an optimal operator is to determine a necessary and sufficient condition for an extremum of the expression $p + \lambda q$. To obtain this condition, we denote by A the optimal operator for which $p + \lambda q$ has an extremal value and by B an arbitrary operator of class R. Then, by substituting in (13)

$$Y=Y(s). \quad Z=AX(t). \quad V=BX(t).$$

we obtain

$$M[Y(s)-AX(t)]M[BX(t)]+\lambda M\{[Y(s)-AX(t)]BX(t)\}=0 \quad (B \in R). \tag{14}$$

For $\lambda = \infty$, equation (14) coincides with the equation obtained by V. S. Pugachev [2].

On the basis of what was proven above for $\lambda \geq 0$ ($\lambda \leq -1$), condition (14), which must hold for any

operator $B \in R$, is necessary and sufficient that operator A realize a minimum (maximum) of $p + \lambda q$ for any $s \in S$. This operator A depends on λ as a parameter.

The greatest and least values of the function $f(p,q)$ are determined as extrema of the function in (5):

$$\Phi(\lambda) = f[p(\lambda),\ q(\lambda)],$$

where $p(\lambda)$ and $q(\lambda)$ are defined by formulae (2).

We now write condition (14) for the particular case when the random function Y(s) is expressed by means of the operator

$$Y(s) = \int_Q \theta(Z(t),\ t,\ s)\,dt, \tag{15}$$

and the nonlinear operator A has the form

$$AX(t) = \int_T \varphi(X(t),\ t,\ s)\,dt, \tag{16}$$

where $\theta(z,t,s)$ is a given function, $\varphi(x,t,s)$ is an arbitrary function and X(t) and Z(t) are random functions.

It is easily seen that the class of operators of the type in (16) is a linear space.

To derive the equation which defines the optimal operator of the form shown in (16), we write the general equation (14) in the form

$$\begin{gathered} M[Y(s) - AX(t)]\,M[BX(u)] + \\ + \lambda\{M[Y(s)\,B_u X(u)] - M[A_t X(t)\,B_u X(u)]\} = 0. \end{gathered} \tag{17}$$

Equation (17) must be satisfied by any operator B of the type given in (16). Consequently, we must set

$$B_u X(u) = \int_T \psi[X(u).\ u,\ s]\,du. \tag{18}$$

where $\psi(x,u,s)$ is an arbitrary function.

By taking (15), (16) and (18) into account, we may write

$$\begin{gathered} M[Y(s) - AX(t)]\,M[B_u X(u)] = \Big[\int_Q\int_{-\infty}^{\infty} \theta(z,\ t,\ s)\,f_z(z,\ t)\,dz\,dt - \\ - \int_T\int_{-\infty}^{\infty} \varphi(x,\ t,\ s)\,f_x(x.\ t)\,dx\,dt\Big]\int_T\int_{-\infty}^{\infty} \psi(x,\ u,\ s)\,f_x(x,\ u)\,dx\,du. \end{gathered} \tag{19}$$

$$M[Y(s)\,B_u X(u)] = \int_Q\int_T du\,dt \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \theta(z,t,\ s)\,\psi(x,\ u,\ s)\,f_{zx}(z,x,\ t,\ u)\,dz\,dx. \tag{20}$$

$$\begin{gathered} M[A_t X(t)\,B_u X(u)] = \int_T\int_T dt\,du \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x,t,s)\,x\times \\ \times\psi(x',\ u,\ s)\,f_{xx}(x,\ x',\ t,\ u)\,dx\,dx'. \end{gathered} \tag{21}$$

where $f_Z(z,t)$ and $f_X(x,t)$ are the one-dimensional probability densities of the random functions Z(t) and Z(t), $f_{ZX}(z,x,t)$ is the two-dimensional joint probability density of the random functions Z(t) and Z(t), and $f_{XX}(x,x',t,t')$ is the two-dimensional probability density of the random function X(t).

If we substitute expressions (19), (20) and (21) into equation (17) and take into account the arbitrariness of the function $\psi(x,u,t)$, we obtain the equation

$$\left[\int\limits_Q dt \int\limits_{-\infty}^{\infty} \theta(z.\ t.\ s)\, f_z(z.\ t)\, dz - \int\limits_T dt \int\limits_{-\infty}^{\infty} \varphi(x.\ t.\ s)\, f_x(x.\ t)\, dx\right] + f_x(x.\ u) +$$

$$+\lambda\left[\int\limits_Q dt \int\limits_{-\infty}^{\infty} \theta(z.\ t.\ s)\, f_{zx}(z.\ x.\ t.\ u)\, dz - \int\limits_T dt \int\limits_{-\infty}^{\infty} \varphi(z.\ t.\ s)\, f_{xx}(z.\ x.\ t.\ u)\, dz\right] = 0 \tag{22}$$

$$(u \in T.\ -\infty < x < \infty).$$

Equation (22) defines the characteristic function, $\varphi(x,t,s)$, of the operator of the form in (16) which realizes the extrema of the expression $p + \lambda q$ when the random function Y(s) has the form given in (15). For $\lambda = \infty$, equation (22) coincides with the equation obtained by Zadeh [6, 7].

We spoke earlier of determining the parameter λ corresponding to the extrema of function $\underline{f}$.

In this paper, we considered the case when all the functions to be investigated are real. There is no particular difficulty in making the transition to the case when the functions are complex.

It should be mentioned in conclusion that the function f(p,q), the criterion for comparing different operators, should be chosen by taking into account the peculiar features of the individual concrete case. This means that, in general, the method of successive approximations will be employed.

SUMMARY

There is deduced the condition of the extremum, of the maximum or minimum values of a certain function $\underline{f}$, of the mean-square error and of the mathematical expectation square of the error of approximating the random function.

The determined general condition is applied to the problem of how to choose the optimum nonlinear integral operator. The paper is based on the ideas and the results of [1-4].

LITERATURE CITED

[1] V. S. Pugachev, "A general condition for the minimum of the mean-square error of dynamic systems," [in Russian]. Automation and Remote Control (USSR) 17, 4 (1956).

[2] V. S. Pugachev, The Theory of Random Functions and Its Application to Problems of Automatic Control, [in Russian], Gostekhizdat (1957).

[3] N. I. Andreev, Definition of an optimal linear dynamic system from an extremum criterion applied to a particular form of functional," Automation and Remote Control (USSR) 18, 7 (1957).

[4] N. I. Andreev, "Regarding the theory of defining optimal dynamic systems," [in Russian], Automation and Remote Control (USSR) 19, 11 (1958).

[5] L. A. Lyusternik and V. I. Sobolev, Elements of Functional Analysis, [in Russian], Gostekhizdat (1951).

[6] L. A. Zadeh, "Optimum nonlinear filters", J. Appl. Phys. vol. 24, No. 4, 1953.

[7] L. A. Zadeh, "A contribution to the theory of nonlinear systems", J. Franklin Inst., vol. 255, No. 5, 1953.

Received November 3, 1958

METHODS OF ANALOG COMPUTER SOLUTION OF LINEAR DIFFERENTIAL EQUATIONS WITH VARIABLE COEFFICIENTS

I. Matyash

(Pardubice , Czechoslovakia)

Two methods are described for the solution of linear differential equations of the type in (1) with variable coefficients by means of analog computers. The methods are illustrated by examples.

INTRODUCTION

In this paper we consider the solution of nonhomogeneous linear differential equations of the form

$$\sum_{i=0}^{n} a_i u^{(i)} = \sum_{i=0}^{m} b_i v^{(i)}, \qquad m \leqslant n. \tag{1}$$

where $a_0, a_1, \ldots, a_n$; $b_0, b_1, \ldots, b_m$ are definite functions of time with derivatives up to the required order, $\underline{v}$ is an arbitrary given function of time such that the right member of the equation is meaningful, and $\underline{u}$ is the solution of equation (1).

Various problems in the domain of control, filtering problems and many others which today attract the attention of numerous specialists, lead to equations of type (1). With the exception of certain special cases, solution of an equation of the type (1) by simple integrations (quadratures) is impossible (cf. [1]), and must be sought by other methods. One of the most acceptable methods is that employing analog computers.

The literature already contains quite detailed descriptions of methods of solving linear differential equations with constant coefficients by means of analog devices (for example, [2,3]). At the same time, block schematics for the analog setups for solving equations with variable coefficients can be found only in individual cases (Cf. [4,5]). For the case when $m = 0$, there are described setups consisting only of integrators, adders and blocks for varying the coefficients. However, in the cases when $m \geq 1$ [when derivatives of the function $v(t)$ appear in the right member of equation (1)], the setups, in the majority of cases also contain differentiators (Cf. [5]) for constructing the required derivatives of the function $v(t)$.

It is well known from the literature on analog devices that differentiators are undesirable as solving elements, since they lead to essential errors. In [2-4] there are described methods for solving equations of the type of (1) with constant coefficients which exclude differentiators from the setups. But, unfortunately, these methods are not applicable to the solution of equations with variable coefficients. In [4], equation (1) is replaced by an equivalent system of first-order equations which is easily solved by means of an analog setup without differentiators.

Below are described two methods for solving differential equations of the form in (1) with variable coefficients, without any necessity of using differentiators in the setup.

The block schematics of the setups consist only of integrators, adders and blocks for varying the coefficients. Figure 1 shows the symbols used for these solving elements. We assume that the transmissions coefficients of the integrators and the adders all equal -1 (this is the case which corresponds to an electronic analog computer).

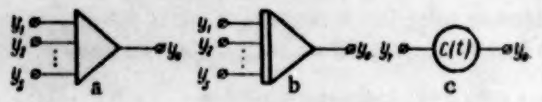


Fig. 1

Then, the input and output quantities of these solving elements satisfy the relationships:

for an adder (Fig. 1,a)

$$y_0(t) = -\sum_{i=1}^{s} y_i(t).$$

for an integrator (Fig. 1,b)

$$y_0'(t) = -\sum_{i=1}^{s} y_i(t).$$

for a block for varying coefficients (Fig. 1, c)

$$y_0(t) = c(t)\, y_1(t).$$

1. First Method

We denote by L and M the linear differential operators:

$$L = \sum_{i=0}^{n} a_i(t) \frac{d^i}{dt^i}. \tag{2}$$

$$M = \sum_{i=0}^{m} b_i(t) \frac{d^i}{dt^i}. \tag{3}$$

Equation (1) can be written in the form

$$L[u] = M[v]. \tag{4}$$

We then define two subsidiary operators:

$$L_1 = \sum_{i=0}^{n} \alpha_i(t) \frac{d^i}{dt^i}. \tag{5}$$

$$M_1 = \sum_{i=0}^{n} \beta_i(t) \frac{d^i}{dt^i}. \tag{6}$$

where $\alpha_i(t)$ and $\beta_i(t)$ $(i = 0, 1, 2, \ldots n)$ are as yet undefined functions of time.

We now fabricate the analog setup (Fig. 2). The quantities V_k $(k = 0, 1, 2, \ldots n-1)$ at the integrator outputs satisfy the equations

$$V_0 = -\alpha_n u - \beta_n v. \tag{7a}$$

$$-V'_{k-1} = V_k + \alpha_{n-k} u + \beta_{n-k} v \quad (k = 1, 2, \ldots, n-1). \tag{7b}$$

$$-V'_{n-1} = \alpha_0 u + \beta_0 v. \tag{7c}$$

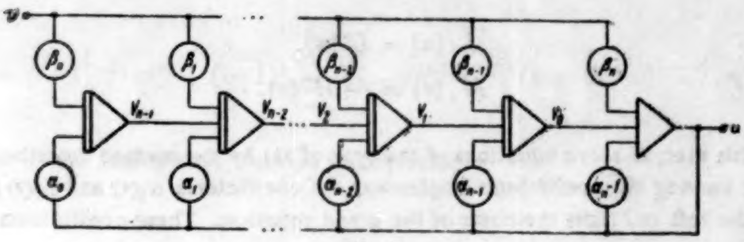


Fig. 2

Consequently,

$$V_k = -\sum_{s=0}^{k} (-1)^{k-s} (a_{n-s} u)^{(k-s)} - \sum_{s=0}^{k} (-1)^{k-s} (\beta_{n-s} v)^{(k-s)}. \tag{8}$$

After substitution in (7b) we obtain, finally,

$$\sum_{j=0}^{n} (-1)^j (\alpha_j u)^{(j)} = -\sum_{j=0}^{n} (-1)^j (\beta_j v)^{(j)}. \tag{9}$$

It follows from this that the scheme of Fig. 2 solves equation (9).

We now denote by L_1^* the operator adjoint to operator L_1:

$$L_1^* [u] = \sum_{i=0}^{n} (-1)^i (\alpha_i u)^{(i)}$$

and by M_1^* the operator adjoint to M_1:

$$M_1^* [v] = \sum_{i=0}^{n} (-1)^i (\beta_i v)^{(i)}.$$

By substitution we obtain equation (9) in the form

$$L_1^* [u] = -M_1^* [v]. \tag{10}$$

By comparing the right and left members of equations (4) and (10) we obtain

$$L_1^* [u] = L [u]. \tag{11}$$

$$-M_1^* [v] = M [v]. \tag{12}$$

By virtue of [1],

$$(K^*)^* = K. \tag{13}$$

where K is any linear differential operator, equations (11) and (12) may be written in the form

$$L_1[u] = L^*[u], \tag{14}$$

$$M_1[v] = -M^*[v]. \tag{15}$$

It follows from this that, to solve equations of the type of (1) by the method described (Fig. 2), it is necessary that the blocks for varying the coefficients implement the coefficients $\alpha_i(t)$ and $\beta_i(t)$ ($i = 0, 1, \ldots, n$) of the adjoint expressions of the left and right members of the given equation. These coefficients are defined by equations (14) and (15).

Since both differential expressions, the original and the one adjoint to it, are both of the same order, we obtain first of all from (15)

$$\beta_{m+k}(t) = 0 \qquad (k = 1, 2, \ldots, n - m) \tag{16}$$

and then

$$\begin{gathered} \beta_m = (-1)^m b_m, \\ \beta_{m-1} = (-1)^m m b_m' + (-1)^{m-1} b_{m-1}, \\ \cdots\cdots\cdots\cdots\cdots \end{gathered} \tag{17}$$

i.e., more generally,

$$\beta_{m-k} = \sum_{i=0}^{k} (-1)^{m-i} \frac{(m-i)!}{(m-k)!(k-i)!} b_{m-i}^{(k-i)} \qquad (k = 0, 1, 2, \ldots, m). \tag{18}$$

In an analogous manner we determine the coefficients α_i ($i = 0, 1, \ldots, n$) from equation (14):

$$\begin{gathered} \alpha_n = (-1)^n a_n, \\ \alpha_{n-1} = (-1)^n n a_n' + (-1)^{n-1} a_{n-1}, \\ \cdots\cdots\cdots\cdots\cdots \end{gathered} \tag{19}$$

and, more generally,

$$\alpha_{n-k} = \sum_{i=0}^{k} (-1)^{n-i} \frac{(n-i)!}{(n-k)!\,(k-i)!} a_{n-i}^{(k-i)} \quad (k = 0, 1, 2, \ldots, n). \tag{20}$$

2. Second Method

We now describe another method for solving equations of the type in (1) by means of analog devices. This method is a generalization of the well-known setup for the case m = 0 (cf. [5]) to the case $0 \le m \le n$. We construct the block schematic for the setup for the method being considered (Fig. 3a or 3b). The quantities V_k ($k = 0, 1, 2, \ldots, n$) in this scheme satisfy the relationships

$$V_0 = -u + \gamma_n v, \tag{21a}$$

$$-V_{k-1}' = V_k + (-1)^{k+1} \gamma_{n-k} v. \tag{21b}$$

Consequently,

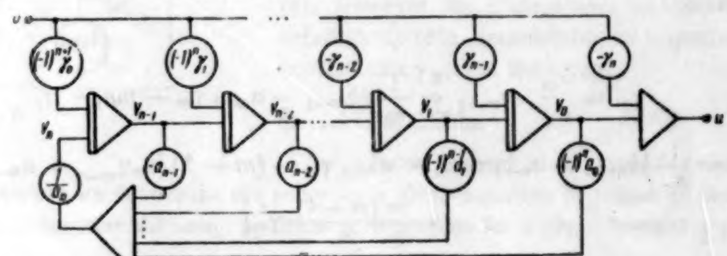
$$V_k = (-1)^{k+1} u^{(k)} + (-1)^k \sum_{s=0}^{k} (\gamma_{n-s} v)^{(k-s)} \qquad (k = 0, 1, 2, \ldots, n). \tag{22}$$

According to Fig. 3, the analog device solves the equation

$$a_n V_n = -\sum_{k=0}^{n-1} (-1)^{n-k} a_k V_k. \tag{23}$$

By substituting from (22) we get

$$\sum_{i=0}^{n} a_i u^{(i)} = \sum_{i=0}^{n} a_i \sum_{s=0}^{i} (\gamma_{n-s} v)^{(i-s)}. \tag{24}$$



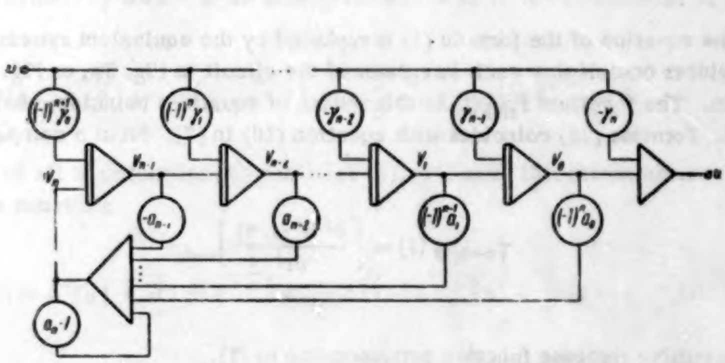


Fig. 3

The left member of equation (24) coincides with the left member of (1). In order that the analog device (with setup as in Fig. 3a or 3b) solve equation (1), it is necessary that the right side of equation (24) also coincide with the right member of equation (1). Thus, we obtain the condition

$$\sum_{i=0}^{n} a_i \sum_{s=0}^{i} (\gamma_{n-s} v)^{(i-s)} = \sum_{i=0}^{m} b_i v^{(i)}. \tag{25}$$

By comparing the expressions for v, v', v",... in the left and right members of equation (25), we obtain for $\gamma_{m+1}, \gamma_{m+2}, \ldots, \gamma_n$

$$\gamma_{m+k}(t) \equiv 0 \qquad (k = 1, 2, \ldots, n-m) \tag{26}$$

and, for $\gamma_1, \gamma_2, \ldots, \gamma_m$, the system of equations

$$a_n\gamma_m = b_m,$$

$$a_n\gamma_{m-1} + a_{n-1}\gamma_m + ma_n\dot{\gamma}_m = b_{m-1}. \tag{27}$$

$$a_n\gamma_{m-2} + a_{n-1}\gamma_{m-1} + a_{n-2}\gamma_m + (m-1)[a_n\gamma_{m-1} + a_{n-1}\dot{\gamma}_m] + \\ + \frac{m(m-1)}{2}\dot{\gamma}_m = b_{m-2}.$$

. . . .

This system may be written in the form

$$b_{m-k} = \sum_{i=0}^{k}\sum_{j=0}^{k-i}\frac{(m-k+i)!}{i!\,(m-k)!}a_{n-j}\gamma^{(i)}_{m-k+i+j} \quad (k=0,1.2.\dots.m). \tag{28}$$

Thanks to the triangular matrix of this system of equations, the coefficients $\gamma_m, \gamma_{m-1},\dots,\gamma_0$ can be computed step by step:

$$\gamma_m = \frac{b_m}{a_n},\ \gamma_{m-1} = \frac{1}{a_n}[b_{m-1} - a_{n-1}\gamma_m - ma_n\gamma'_m],$$

$$\gamma_{m-2} = \frac{1}{a_n}\Big[b_{m-2} - a_{n-1}\gamma_{m-1} - a_{n-2}\gamma_m - (m-1)(a_n\gamma'_{m-1} + a_{n-1}\gamma'_m) - \\ - \frac{m(m-1)}{2}\gamma'_m\Big] \tag{29}$$

. . . .

In [4] (§5, 2), the equation of the form in (1) is replaced by the equivalent system of first-order equations. One easily convinces oneself that each integrator of the circuit in Fig. 3a, or Fig. 3b solves one of the equations of this system. The functions $F_{n-k}(t)$ in this system of equations coincide, obviously, with the variable coefficients $\gamma_k(t)$. Formula (28) coincides with equation (16) in [7]. From a comparison, we obtain the important relationship

$$\gamma_{n-1-j}(t) = \left[\frac{\partial^j W(t,\tau)}{\partial t^j}\right]_{\tau=t}, \tag{30}$$

where $W(t,\tau)$ is the impulsive response function corresponding to (1).

In solving the problem of optimal filtering, and in analogous cases, the problem can frequently arise of implementing a linear system with variable coefficients in accordance with a given impulsive response function. Works [6, 7] are devoted to the problem of determining the coefficients of a differential equation from the impulsive response function corresponding to it.

It is clear from equation (30) that, for the block schematic of a setup by this method, it is not necessary to compute the coefficients b_i ($i = 0,1,\dots,m$) of the right side of equation (1), but only the coefficients a_i ($i = 0,1,\dots,n$) of the left side, which makes the task easier. The coefficients γ_i ($i = 0,1,\dots,n$), required for solving a problem by this method, are obtained from equation (30) by differentiating the impulsive response function.

Obviously, the order of the right side of the equation, corresponding to the given impulsive response function, determines the number of zero derivatives in the right side of (30). According to [7], it follows from (26), (29) and (30) that the first nonzero derivative will be the j'th derivative, where $j = n-m-1$.

If $m = n$, then $j = -1$. In this case, the function $W(t,\tau)$ contains the term $\gamma_n(\gamma)\,\delta(t-\tau)$, where δ is the unit pulse function.

But in the majority of practical cases, $m < n$. In these cases, it is possible to construct the first of the

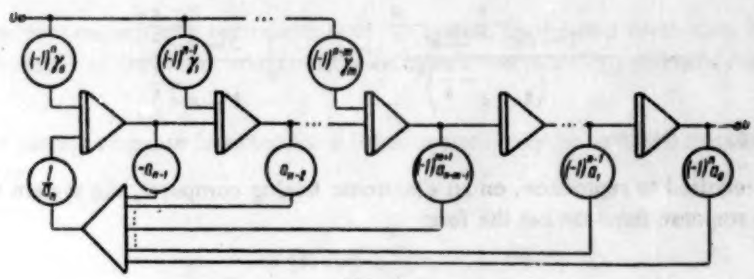


Fig. 4

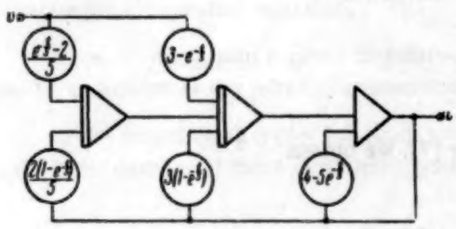


Fig. 5

setups cited above (Fig. 4). In case m = 0, this circuit, obviously, coincides with a well-known circuit (Cf. [5]). By the same method it is also possible to construct the scheme shown in Fig. 3b.

The method described is a variation of the method in [4]. However, the present work includes the derivation of relationship (30), determining an important property of the coefficients $\gamma_k(t)$ of the setup.

3. Examples.

We now present an example for each of the methods described. In the first example we determine the setup for a given equation by means of the first method, and in the second example we determine the setup by the second method for a given impulsive response function.

Example 1. It is required, by means of an analog device, to solve the differential equation

$$5\left(1-e^{-\frac{t}{5}}\right)u'' + \left(3-e^{-\frac{t}{5}}\right)u' + \frac{2}{5}u = \left(3-e^{-\frac{t}{5}}\right)v' + \frac{2}{5}v.$$

It is necessary first of all to determine, by formulae (4), (14) and (15), the expressions adjoint to the left and right members of this equation:

$$L_1[u] = L^*[u] = 5\left(1-e^{-\frac{t}{5}}\right)u'' - 3\left(1-e^{-\frac{t}{5}}\right)u' + \frac{2}{5}\left(1-e^{-\frac{t}{5}}\right)u$$

$$-M_1[v] = M^*[v] = -\left(3-e^{-\frac{t}{5}}\right)v' + \frac{1}{5}\left(2-e^{-\frac{t}{5}}\right)v.$$

From this we determine the coefficients α_i and β_i:

$$\alpha_0 = \frac{2}{5}\left(1-e^{-\frac{t}{5}}\right), \quad \alpha_1 = -3\left(1-e^{-\frac{t}{5}}\right) \quad \alpha_2 = 5\left(1-e^{-\frac{t}{5}}\right).$$

$$\beta_0 = -\frac{1}{5}\left(2-e^{-\frac{t}{5}}\right), \quad \beta_1 = 3-e^{-\frac{t}{5}}.$$

From Fig. 2 we obtain the block schematic for solving the given problem (Fig. 5). This block schematic can be formed by different methods and from these similar circuits can also be found. In order to realize the blocks for varying the coefficients by means of multipliers with servo systems, the most advantageous circuit is the one shown in Fig. 6. The constant coefficients in this scheme can be implemented by means of the transmission coefficients of adders and integrators. The multiplication of the quantities u(t) and v(t) by the variable coefficient $1-e^{-t/5}$ can be implemented by one multiplier with two potentiometers.

The second method would not work in this case since the coefficients

$$\gamma_0 = -\frac{1}{25}\,\frac{7 - 6e^{-\frac{t}{5}} + e^{-\frac{2t}{5}}}{\left(1 - e^{-\frac{t}{5}}\right)^2}, \quad \gamma_1 = \frac{1}{5}\,\frac{3 - e^{-\frac{t}{5}}}{1 - e^{-\frac{t}{5}}}$$

tend to infinity for $t \to 0$.

Example 2. It is required to reproduce, on an electronic analog computer, the system with variable parameters whose impulsive response function has the form

$$W(t, \tau) = t^2 + \tau^2 e^{\tau} e^{-t} - 2\tau t.$$

Obviously, $n = 3$. The fundamental system of the homogeneous equation is

$$u_1 = t, \quad u_2 = t^2, \quad u_3 = e^{-t}.$$

For the coefficients of the left member of the equation, by [6] or [7], we obtain

$$a_0 = 2, \quad a_1 = -2t, \quad a_2 = t^2, \quad a_3 = t^2 + 2t + 2.$$

The coefficients $\gamma_0, \gamma_1, \ldots$ we determine in accordance with equation (30):

$$\gamma_2 = W(t, t) = 0, \quad \gamma_1 = \left[\frac{\partial W(t, \tau)}{\partial t}\right]_{\tau = t} = -t^2, \quad \gamma_0 = \left[\frac{\partial^2 W(t, \tau)}{\partial t^2}\right]_{\tau = t} = t^2 + 2.$$

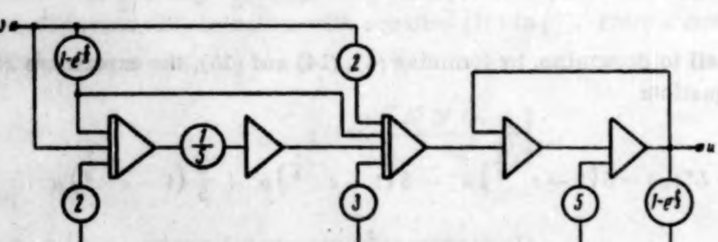


Fig. 6

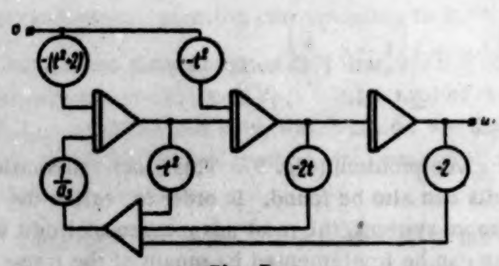


Fig. 7.

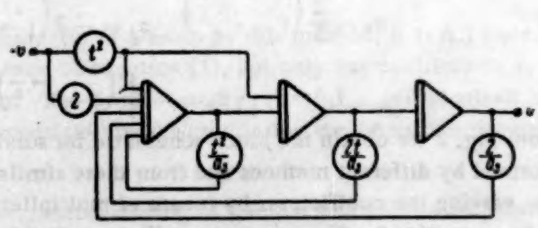


Fig. 8

The block schematic of the system with the given impulsive response function is shown in Fig. 7. We obtain a setup similar to the one we had in example 1 (Fig. 8). Four multipliers are required to implement this scheme.

One easily convinces oneself that reproduction of the system by the first method is, in this case, less satisfactory, due to the necessity of including adders in the setup and the more complicated realization of the variable coefficients.

Note. Every impulsive response function for a linear system may be written in the form of a sum of products (Cf. [7]):

$$W(t,\tau)=\sum_{i=1}^{q} f_i(t)\,g_i(\tau).$$

Each term $f_i(t)g_i(\tau)$ ($i = 1,2,\ldots,q$) can be considered as the impulsive response function corresponding to a first-order differential equation.

Thus, starting from a given impulsive response function $W(t,\tau)$, a system with variable parameters may also be presented as the parallel connection of q first-order systems (Cf. [6]).

If the functions $f_i(t)$ ($i = 1,2,\ldots,q$) and the functions $g_i(\tau)$ ($i = 1,2,\ldots,q$) are linearly independent, then q is the least number of these first-order systems.

SUMMARY

This work described two methods of solving linear differential equations of the form of (1) with variable coefficients by means of analog devices, where no differentiators were used in the setups.

With zero initial conditions on the integrators, we obtain, at the scheme's output, the solution of equation (1) defined by the expression

$$u(t)=\int_{-\infty}^{t} W(t,\tau)\,v(\tau)\,d\tau.$$

We obtain any other solution when the initial conditions on the integrators are not zero (this is proven in [4]).

From the point of view of the number of operations necessary to compute the coefficients which must be realized, the first method will be better in the majority of cases, when it is necessary to implement a system with variable parameters for a given equation, and the second method will be better when a system must be reproduced from a given impulsive response function.

LITERATURE CITED

[1] V. V. Stepanov, Course in Differential Equations [in Russian], Gostekhizdat (1953).

[2] C. L. Johnson, Analog Computer Techniques McGraw-Hill, 1956.

[3] J. Matyas, "Programming a linear differential equation with constant coefficients on a differential analyzer," Slaboproudy obzor (in the press).

[4] J. H. Laning and R. H. Battin, Random Processes in Automatic Control, McGraw-Hill, 1956.

[5] A. V. Solodov, "Statistical investigation of nonstationary processes in linear systems by the use of inverse analog devices," [in Russian], Automation and Remote Control (USSR) 19, 4 (1958).

[6] A. M. Batkov, "On the question of synthesizing dynamic systems with variable parameters," [in Russian]. Automation and Remote Control (USSR) 19, 1 (1958).

[7] V. Borskii, "On the properties of the impulsive response functions of systems with variable parameters," [in Russian], Automation and Remote Control (USSR) 20, 6 (1959).

Received August 18, 1958

ON THE PROPERTIES OF THE IMPULSIVE RESPONSE FUNCTION OF SYSTEMS WITH VARIABLE PARAMETERS

V. Borskii

(Pardubitse, ChSR)

A definite class of functions of two variables is considered as to the use of their properties for solving the problem of determining a linear differential equation from a given impulsive response function, and conversely.

In many cases of solving automatic control problems, the necessity arises of investigating systems with variable parameters. A linear system whose parameters are functions of time is described by a linear differential equation with variable coefficients. The analysis of such systems is the main topic of the works of L. A. Zadeh [1-6].

The behavior of physical systems as functions of their input excitations can be expressed by means of their impulsive response functions. In certain cases, for example, in the synthesis of optimal filters, the impulsive response is given and it is necessary to determine the equation corresponding to it. A similar problem was solved by A. M. Batkov [7]. However, the method described by him for determining the right member of the equation sought is attended with great difficulty. Moreover, in formulae (4) of that paper, giving the initial conditions, there are essential errors (cf. Corollary 2 to Theorem 4 of the present paper) and the assertion that Formula (6) is a solution of Equation (5) is invalid.

In the present work are presented formulae which allow, with comparative facility, the differential equation to be determined from a given impulsive response. Also obtained are general properties of the impulsive responses which, for greatest clarity, are given in the form of theorems.

Properties of the Impulsive Responses

Let there be given a function of two variables, $W(t,\tau)$, continuous in the interval considered and with continuous partial derivatives of all orders which enter into our formulae.

We use the notation

$$\begin{aligned} W_{00}(t, \tau) &= W(t, \tau), \\ W_{i0}(t, \tau) &= \frac{\partial^i}{\partial t^i} W(t, \tau), \\ W_{0j}(t, \tau) &= \frac{\partial^j}{\partial \tau^j} W(t, \tau), \\ W_{00}^{(k)}(t, \tau) &= \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \tau}\right)^k W(t, \tau), \\ W_{ij}^{(k)}(t, \tau) &= \frac{\partial^i}{\partial t^i}\frac{\partial^j}{\partial \tau^j}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \tau}\right)^k W(t, \tau). \end{aligned} \tag{1}$$

Definition 1. A function of two variables, $W(t,\tau)$, will be called an n'th-order degenerate function if t there exists an integer, n, such that

$$\begin{vmatrix} W_{00}(t,\tau) & W_{01}(t,\tau) & \dots W_{0(n-1)}(t,\tau) \\ W_{10}(t,\tau) & W_{11}(t,\tau) & \dots W_{1(n-1)}(t,\tau) \\ \vdots & \vdots & \vdots \\ W_{(n-1)0}(t,\tau) & W_{(n-1)1}(t,\tau) & \dots W_{(n-1)(n-1)}(t,\tau) \end{vmatrix} \neq 0. \tag{2}$$

but

$$\begin{vmatrix} W_{00}(t,\tau) & W_{01}(t,\tau) & \dots W_{0n}(t,\tau) \\ W_{10}(t,\tau) & W_{11}(t,\tau) & \dots W_{1n}(t,\tau) \\ \vdots & \vdots & \vdots \\ W_{n0}(t,\tau) & W_{n1}(t,\tau) & \dots W_{nn}(t,\tau) \end{vmatrix} = 0. \tag{3}$$

lin ind.

In the present work, the function $W(t,\tau)$ will always be considered to be n'th-order degenerate.

Theorem 1. There exists one, and only one, homogeneous linear n'th-order differential equation

$$\sum_{i=0}^{n} a_i(t)\, u^{(i)}(t) = 0 \tag{4}$$

such that the functions $W_{0j}(t,\tau)$ ($j = 0,1,\dots,n-1$) constitute a fundamental system of solutions for any value of the variable τ, considered as a parameter.

In view of the fact that, by virtue of relationship (3), the functions $W_{i0}(t,\tau)$ ($i = 0,1,\dots,n$) of the argument τ, for any t, are independent, there exist continuous functions $a_i(t)$ ($i = 0,1,\dots,n$) such that

$$\sum_{i=0}^{n} a_i(t)\, W_{ij}(t,\tau) = 0 \qquad (j = 0,1,\dots,n). \tag{5}$$

These relationships are a homogeneous system of algebraic equations for the functions $a_i(t)$ which, in view of (2), have a unique (linearly independent) fundamental [8] solution.

In accordance with (2), the functions $W_{0j}(t,\tau)$ ($j = 0,1,\dots,n$) and τ, for any t, are independent, and may be considered as a fundamental system of solutions of equation (4).

Corollary to theorem 1. Each n'th-order degenerate function of two variables, as a function of t for any τ, is the solution of an n'th-order linear differential equation whose coefficients do not depend on τ. This equation is unique.

Theorem 2. Each n'th-order degenerate function of two variables consists of a sum of products

$$W(t,\tau) = \sum_{j=1}^{n} u_j(t)\, \varphi_j(\tau),$$

where the $u_j(t)$ ($j = 1,2,\dots,n$) are linearly independent functions of t and the $\varphi_j(\tau)$ ($j = 1,2,\dots,n$) are linearly independent functions of τ.

Since, according to the corollary to theorem 1, the function $W(t,\tau)$ is a solution of the homogeneous equation (4), it may be expressed by means of any fundamental system of solutions of this equation [9]. Let $u_j(t)$ ($j = 1,2,\dots,n$) be a fundamental set of solutions of equation (4). We determine the functions $\varphi_j(\tau)$, satisfying equation (6), from the system of equations

$$\sum_{j=1}^{n} u_j^{(i)}(t)\varphi_j(\tau) = W_{i0}(t, \tau) \quad (i = 0, 1, \ldots, n-1). \tag{7}$$

which, by virtue of the independence of the functions $u_j(t)$, has a unique solution. In (2), the functions $W_{i0}(t,\tau)$ ($i = 0,1,\ldots,n-1$) are independent, so that it is obvious from equation (7) that the functions $\varphi_j(\tau)$ ($j = 1,2,\ldots,n$) are also independent.

__Theorem 3.__ Any function of two variables of the form

$$W(t,\tau) = \sum_{i=1}^{q} f_i(t)\, q_i(\tau) \tag{8}$$

is degenerate.

The order of degeneracy, n, of the function $W(t,\tau)$ equals q ($n = q$) in those cases when the functions $f_i(t)$ ($i = 1,2,\ldots,q$) and the functions $q_i(\tau)$ ($i = 1,2,\ldots,q$) are independent; in the remaining cases, $n < q$.

The proof follows from definition 1.

__Theorem 4.__ Let there be given a linear differential equation

$$\sum_{i=0}^{n} a_i(t)\, u^{(i)}(t) = \sum_{j=0}^{m} b_j(t)\, v^{(j)}(t) \quad (m < n). \tag{9}$$

where $a_i(t)$ and $b_j(t)$ are continuous functions in the interval considered and have continuous derivatives up to the required order and $v(t)$ is any function for which the right member of equation (9) is meaningful. Then, there exists one, and only one, function of two variables, $W(t,\tau)$, such that

$$u(t) = \int_{-\infty}^{t} W(t,\tau)\, v(\tau)\, d\tau \tag{10}$$

is a solution of this differential equation.

We first differentiate equation (10) $\underline{n}$ times with respect to $\underline{t}$:

$$\begin{aligned} u'(v) &= \int_{-\infty}^{t} W_{10}(t,\tau)\, v(\tau)\, d\tau + c_{11}(t)\, v(t). \\ u''(t) &= \int_{-\infty}^{t} W_{20}(t,\tau)\, v(\tau)\, d\tau + c_{12}(t)\, v(t) + c_{22}(t)\, v'(t). \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ u^{(\nu)}(t) &= \int_{-\infty}^{t} W_{\nu 0}(t,\tau)\, v(\tau)\, d\tau + \sum_{\mu=1}^{\nu} c_{\mu\nu}(t)\, v^{(\mu-1)}(t). \end{aligned} \tag{11}$$

The functions $c_{\mu\nu}(t)$ are defined by the recursion formulae

$$c_{\mu\nu}(t) = 0 \qquad \text{for } \mu > \nu. \tag{12a}$$

$$c_{\mu\nu}(t) = W_{00}(t,t) \qquad \text{for } \mu = \nu. \tag{12b}$$

$$c_{\mu\nu}(t) = \frac{d}{dt} c_{\mu(\nu-1)}(t) + C_{\nu-1}^{\mu-1} W_{00}^{(\nu-\mu)}(t, t) \quad \text{for } \mu < \nu. \tag{12c}$$

or, more generally,

$$c_{\mu\nu}(t) = \sum_{\lambda=\mu}^{\nu} C_{\lambda-1}^{\mu-1} W_{(\nu-\lambda)0}^{(\lambda-\mu)}(t, t). \tag{13}$$

where $C_l^k = \frac{l\,!}{k!\,(l-k)!}$ is a binominal coefficient (cf. also formula (30) in the Appendix).

By substituting expressions (11) in the left member of equation (9) we obtain

$$\int_{-\infty}^{t} \sum_{i=0}^{n} a_i(t) W_{i0}(t, \tau) v(\tau) d\tau + \sum_{i=0}^{n} \sum_{\mu=1}^{i} a_i(t) c_{\mu i}(t) v^{(\mu-1)}(t) = \sum_{j=0} b_j(t) v^{(j)}(t). \tag{14}$$

We assume that $W(t,\tau)$ is a solution of the homogeneous equation

$$\sum_{i=0}^{n} a_i(t) W_{i0}(t, \tau) = 0. \tag{15}$$

This assumption corresponds to the physical assertion that $W(t,\tau)$ is an impulsive response or, in other words, $W(t,\tau)$ defines the behavior of a linear dynamic system as a function of a unit input pulse [1, 2].

By substituting (15) in equation (14) we obtain, in accordance with formula (12a),

$$\sum_{\mu=1}^{n} \sum_{i=\mu}^{n} a_i(t) c_{\mu i}(t) v^{(\mu-1)} = \sum_{j=0}^{m} b_j(t) v^{(j)}(t).$$

If we set $i = \nu$, $j = \mu - 1$ and carry out a comparison of coefficients of the derivatives of the function $v(t)$, we get

$$\sum_{\nu=\mu}^{n} a_\nu(t) c_{\mu\nu}(t) = b_{(\mu-1)}(t) \quad (\mu = 1, 2, \ldots, n). \tag{16}$$

The function $W(t,\tau)$ exists and is completely defined by conditions (15) and (16).

That it is unique can be proved along the same lines as were followed for the case $b_0 = 1$, $b_l = 0$ ($l = 1,2,\ldots,m$) in [12].

Corollary 1 of theorem 4. The general solution of equation (9) has the form

$$u(t) = f(t) + \int_{-\infty}^{t} W(t, \tau) v(\tau) d\tau.$$

where $f(t)$ is the general solution of the homogeneous equation (4).

Corollary 2 of theorem 4. If $v(t) = 0$ for $t < t_0$ and the initial values of this function and its derivatives in approaching the point t_0 from the right equal $v^{(i)}(t_0^+)$ ($i = 0,1,2,\ldots$), then initial conditions for solution (10) of differential equation (9) are defined by the expressions

$$\mu(t_0^+) = 0,$$
$$u^{(\nu)}(t_0^+) = \sum_{\mu=1}^{\nu} c_{\mu\nu}(t_0)\, v^{(\mu-1)}(t_0^+) \qquad (\nu = 1, 2 \ldots). \tag{17}$$

Definition 2. The function of two variables, defined by theorem 4 for the linear differential equation (9), describing a linear dynamic system, we shall call the impulsive response of this system.

Theorem 5. The impulsive response corresponding to differential equation (9) is a degenerate function.

The impulsive response must be a solution to the homogeneous equation (15),and, consequently, it is possible to express it as the sum of products as in (6) (cf. the proof of theorem 2). We note that, in this case, the functions $\varphi_j(\tau)$ must not satisfy the conditions of independence. Each sum of a finite number of products of the form of (6) is a degenerate function by virtue of theorem 3.

Theorem 6. There exists one, and only one, n'th-order linear differential equation for which the given n'th-order degenerate function of two variables, $W(t,\tau)$, is theimpulsive response.

In order that the function $W(t,\tau)$ be an impulsive response it must satisfy conditions (15) and (16). According to the corollary to theorem 1, the homogeneous equation (15) is uniquely defined and thus the left member of the equation sought will also be uniquely determined. The coefficients of the right member of this equation are uniquely determined by formula (16) or by the formulae of the Appendix.

Theorem 7. Each n'th-order degnerate function of two variables, $W(t,\tau)$, is the impulsive response of an infinite number of linear differential equations of order q, $q > n$.

A degenerate function of order n can be expressed by the formula

$$W(t,\tau) = \sum_{j=1}^{q} \bar{u}_j(t)\, \bar{\varphi}_j(\tau) \qquad (q > n),$$

where $\bar{u}_j(t)$ $(j = 1,2,\ldots,q)$ are linearly independent functions and $\bar{\varphi}_j(\tau)$ $(j = 1,2,\ldots,q)$ are linearly dependent functions of which n functions are independent (the Wronskian, $W_{[\bar{\varphi}(\tau),\, \bar{\varphi}_2(\tau),\ldots,\, \bar{\varphi}_q(\tau)]}$ has rank n).

By virtue of their independence, the functions $\bar{u}_j(t)$ $(j = 1,2,\ldots,q)$ may be considered as a fundamental system of solutions of the q'th-order equation, and their coefficients, $\bar{a}_j(t)$, may be determined by standard methods [9].

The coefficients $\bar{b}_j(t)$ of the right member of the desired equation are obtained from relationship (16) for $\mu = 1,2,\ldots,q$.

These equations of orders $q > n$ may also be obtained by multiplying the left and right members of the n'th-order equation, determined by theorem 6, by any linear differential operator.

Theorem 8. If $W_{i0}(t,t) = 0$ for $i = 0,1,\ldots,(l-1)$, but $W_{l0}(t,t) \neq 0$, then the order, m, of the right side of the equation corresponding to the given impulsive response $W(t,\tau)$ is obtained from the formula

$$m = q - l - 1,$$

where q is the order of the left member of this equation.

The proof follows from equation (16) when relationship (12), or (13), is taken into account.

Theorem 9. The impulsive response $G(t,\tau)$, corresponding to the differential equation

$$\sum_{i=0}^{n} a_i(t)\, u^{(i)}(t) = a_n(t)\, v(t), \tag{18}$$

is determined by the expression

$$G(t,\tau)=\sum_{i=1}^{n}\frac{\partial \ln W[u_i(\tau)]}{\partial u_i^{(n-1)}(\tau)}u_i(t)=\frac{\begin{vmatrix} u_1(\tau) & u_2(\tau) & \dots & u_n(\tau)\\ u_1'(\tau) & u_2'(\tau) & \dots & u_n'(\tau)\\ \vdots & \vdots & & \vdots\\ u_1^{(n-2)}(\tau) & u_2^{(n-2)}(\tau) & \dots & u_n^{(n-2)}(\tau)\\ u_1(t) & u_2(t) & \dots & u_n(t)\end{vmatrix}}{\begin{vmatrix} u_1(\tau) & u_1(\tau) & \dots & u_n(\tau)\\ u_1'(\tau) & u_2'(\tau) & \dots & u_n(\tau)\\ \vdots & \vdots & & \vdots\\ u_1^{(n-1)}(\tau) & u_2^{(n-1)}(\tau) & \dots & u_n^{(n-1)}(\tau)\end{vmatrix}} \tag{19}$$

where the $u_i(t)$ form a fundamental system of solutions of the corresponding homogeneous differential equation.

As was shown in [9] (this can be verified by theorem 4), the impulsive response, defined in definition 2, for equation (18) must be a solution of the corresponding homogeneous equation, and must satisfy the following conditions:

$$\left[\frac{\partial^i}{\partial t^i}G(t,\tau)\right]_{\tau=t}=0 \qquad (i=0,1,\dots,n-2), \tag{20}$$

$$\left[\frac{\partial^{(n-1)}}{\partial t^{(n-1)}}G(t,\tau)\right]_{\tau=t}=1. \tag{21}$$

By the rules for differentiating determinants, it is easily verified that $G(t,\tau)$ satisfies all these conditions.

<u>Theorem 10.</u> The function of two variables

$$W(t,\tau)=\sum_{j=0}^{n-1}\left(-\frac{\partial}{\partial\tau}\right)^j\left[\frac{b_j(\tau)}{a_n(\tau)}G(t,\tau)\right], \tag{22}$$

where $G(t,\tau)$ is defined by theorem 9, is the impulsive response corresponding to equation (9).

This theorem is proven in [2] by the use of the unit pulse function, and in [10] by means of integration by parts.

<u>Example.</u> We consider an application of the theorems given for the determination of a differential equation corresponding to a given impulsive response $W(t,\tau)$.

Let

$$W(t,\tau)=\frac{4e^{0.4(\tau-t)}-e^{0.2(\tau-t)}-e^{0.2\tau-0.4t}}{5(1-e^{-0.2t})}. \tag{23}$$

This function can be expressed as the sum of two products:

$$W(t,\tau)=\varphi_1(\tau)u_1(t)+\varphi_2(\tau)u_2(t), \tag{24}$$

letting, for example,

$$\varphi_1(\tau)=e^{0.4\tau},\quad \varphi_2(\tau)=-e^{0.2\tau},\quad u_1(t)=\frac{4e^{0.4t}}{5(1-e^{-0.2t})},\quad u_2(t)=\frac{e^{-0.4t}+e^{-0.2t}}{5(1-e^{-0.2t})}. \tag{25}$$

Since the functions $u_1(t)$, $u_2(t)$, $\varphi_1(\tau)$ and $\varphi_2(\tau)$ are independent, as may be easily verified then, according to theorems 3 and 6, the equation sought is a second-order equation. The functions $u_1(t)$ and $u_2(t)$ are a fundamental system of solutions of the homogeneous equation corresponding to the equation sought. Therefore,

$$a_2=25(1-e^{-0.2t}),\quad a_1=5(3-e^{-0.2t}),\quad a_0=2. \tag{26}$$

By formula (12) or (13), or by the formulae of the Appendix, we determine the coefficient $c_{\mu\nu}(t)$, taking into account that

$$W_{00}(t,t)=\frac{3-e^{-0.2t}}{5(1-e^{-0.2t})},$$

$$W'_{00}(t,t)=-\frac{2}{25}\frac{e^{-0.2t}}{(1-e^{-0.2t})^2}, \tag{27}$$

$$W_{10}(t,t)=-\frac{1}{25}\frac{7-6e^{-0.2t}+e^{-0.4t}}{(1-e^{-0.2t})^2}.$$

By virtue of relationship (16), we can determine the coefficients of the right member of the equation sought:

$$b_0(t)=2,\quad b_1(t)=5(3-e^{-0.2t}). \tag{28}$$

Consequently, to the given impulsive response of (23) there corresponds the differential equation

$$25(1-e^{-0.2t})u''(t)+5(3-e^{-0.2t})u'(t)+2u(t)=5(3-e^{-0.2t})v'(t)+2v(t). \tag{29}$$

SUMMARY

The principal aim of this paper was the finding of a differential equation from the given impulsive response which corresponds to it. It was established that this problem is not unique, and relationships were derived for determining completely the equation of lowest order which satisfies the conditions set. By the use of equations of higher order it is possible, in individual cases, to get to an equation whose coefficients are more advantageous from the point of view of their programming by means of an analog device – a differential analyzer [11].

If it is required to find the impulsive response corresponding to a given differential equation, this may be completely and uniquely determined by means of theorem 4, but is more conveniently done by means of theorems 9 and 10.

All assertions proven remain valid if the variables t and τ are interchanged.

Analogously to the manner in which the relationships for equation (9) were derived for $m < n$, one may derive the relationships also for the case $m = n$, setting, for example,

$$u(t)=\alpha(t)v(t)+\int_{-\infty}^{t}\overline{W}(t,\tau)v(\tau)d\tau.$$

For $m > n$, a similar expression should be used, with $m-n$ derivatives of the function $v(t)$ added.

The author wishes to thank J. Matyash for many good counsel and significant help in the formulation and proof of the theorems.

Appendix

1. For determining the coefficients $c_{\mu\nu}(t)$, it is convenient to use the following matrix product:

$$\left\|\begin{matrix} c_{11}(t) & c_{12}(t) & c_{13}(t) & c_{14}(t) & c_{15}(t) & \ldots \\ c_{22}(t) & c_{23}(t) & c_{24}(t) & c_{25}(t) & \ldots & \ldots \\ c_{33}(t) & c_{34}(t) & c_{35}(t) & \ldots & \ldots & \ldots \\ c_{44}(t) & c_{45}(t) & \ldots & \ldots & \ldots & \ldots \\ c_{55}(t) & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{matrix}\right\| =$$

$$= \left\|\begin{matrix} 1 & 1 & 1 & 1 & 1 & \ldots \\ 1 & 2 & 3 & 4 & 5 & \ldots \\ 1 & 3 & 6 & 10 & \ldots & \ldots \\ 1 & 4 & 10 & \ldots & \ldots & \ldots \\ 1 & 5 & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{matrix}\right\| \times \left\|\begin{matrix} W_{00}(t,t) & W_{10}(t,t) & W_{20}(t,t) & W_{30}(t,t) & \ldots \\ 0 & W'_{00}(t,t) & W'_{10}(t,t) & W'_{20}(t,t) & \ldots \\ 0 & 0 & W''_{00}(t,t) & W''_{10}(t,t) & \ldots \\ 0 & 0 & 0 & W'''_{00}(t,t) & \ldots \\ 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{matrix}\right\| \tag{30}$$

2. In determining the coefficients of the right member of equation (9), it is possible to present relationship (16) in the more graphic forms:

a) for a first-order equation

$$b_0(t) = a_1(t) W_{00}(t, t); \tag{31}$$

b) for a second-order equation

$$\begin{aligned} b_0(t) &= a_1(t) W_{00}(t,t) + a_2(t)[W'_{00}(t, t) + W_{10}(t, t)], \\ b_1(t) &= a_2(t) W_{00}(t, t); \end{aligned} \tag{32}$$

c) for a third-order equation

$$\begin{aligned} b_0(t) &= a_1(t) W_{00}(t, t) + a_2(t)[W'_{00}(t, t) + W_{10}(t, t)] + \\ &+ a_3(t)[W''_{00}(t, t) + W'_{10}(t, t) + W_{20}(t, t)], \\ b_1(t) &= a_2(t) W_{00}(t, t) + a_3(t)[2W'_{00}(t, t) + W_{10}(t, t)], \\ b_2(t) &= a_3(t) W_{00}(t, t). \end{aligned}$$

LITERATURE CITED

[1] L. A. Zadeh, Frequency Analysis of Variable Networks, Proc. IRE, vol. 38, No. 3, 1950.

[2] L. A. Zadeh, The Determination of the Impulse Response of Variable Networks, J. of Applied Physics, vol. 21, No. 6, 1950.

[3] L. A. Zadeh, Circuit Analysis of Linear Varying-Parameter Networks, J. of Applied Physics, vol. 21, No. 11, 1950.

[4] L. A. Zadeh, On Stability of Linear Varying-Parameter Systems, J. of Applied Physics, vol. 22, No. 4, 1951.

[5] L. A. Zadeh, Initial Conditions in Linear Varying-Parameter Systems, J. of Applied Physics, vol. 22, No. 6, 1951.

[6] L. A. Zadeh, A General Theory of Linear Signal Transmission Systems, J. of Franklin Inst., IV, 1952.

[7] A. M. Batkov, "On the question of synthesizing dynamic systems with variable parameters," [in Russian], Automation and Remote Control (USSR) 19, 1 (1958).

[8] I. N. Bronshtein and K. A. Semendyaev, Handbook of Mathematics, [in Russian], Gostekhizdat,(Moscow, 1956).

[9] V. V. Stepanov, Course in Differential Equations, [in Russian]. Gostekhizdat (Moscow, 1953).

[10] J. H. Laning and R. H. Battin, Random Processes in Automatic Control, N.-Y., McGraw-Hill, 1956.

[11] J. Matyash, "Methods of analog computer solution of linear differential equations with variable coefficients," [in Russian]. Automation and Remote Control (USSR) 20, 7 (1959).

[12] K. S. Miller, "The one- sided Green's function" of Applied Physics, vol. 22, No. 8, 1951.

Received February 1, 1959

A CRITERION OF CONTROL INACCURACY

Kh. V. Ruubel*

(Tallin)

A theoretical basis is adduced for a criterion of control inaccuracy which allows objective comparisons to be made between different variants of automatic control systems.

1. Preliminary Remarks.

Almost every automatic control problem leads to the task of minimizing * the average value of some composite technicoeconomic parameter over the entire period T of use, i.e, to the minimization of the quantity

$$\overline{P} = \frac{1}{T}\int_0^T P dt. \quad (1)$$

The parameter P whose mean value $\overline{P}$ is required to be minimized is conventionally called the "importance" [1].

Let us consider some examples.

Example 1. Controlling the CO_2 content of flue gases of a boiler aggregate [2]. The deviation of the controlled quantity $\underline{x}$ is here the deviation of the CO_2 content from a given optimal value. The importance is expressed by the total power loss (the total of the heat loss from the escaping gases and the heat loss from the chemically incomplete combustion).

As the CO_2 content increases, the heat loss from the escaping gases decreases but the heat loss due to incomplete combustion increases. With the optimal CO_2 content, the total power loss will have its minimum value. A deviation of the CO_2 content from the optimum value on either side will lead to an increase in the total power loss. The graph of the function P = P(x) is given on Fig. 1.

Example 2. Controlling the water level in a boiler aggregate's drums [3]. The deviation of the controlled quantity $\underline{x}$ is the deviation of the water level from the given normal value. The importance P is the cost related to the amortization of the boiler due to a shortening of its life of service, and also to other expenditures due to the deviation of the level from its normal value at some given moment of use. At the normal level, the cost is minimal. With increasing deviation, the cost increases. The graph of the function P = P(x) is shown in Fig. 2.

In the majority of cases, the relationship between the deviations of the controlled quantity and the importance P is not constant, but is some function of various extraneous arguments (for example, in example 1 this

* If it is required to maximize the time average of some parameter, we can reduce this case to a minimization problem by subtracting the given parameter from a constant and considering the difference as the parameter P.

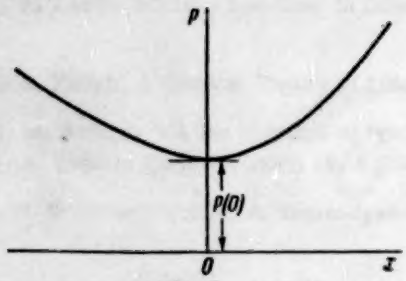


Fig. 1

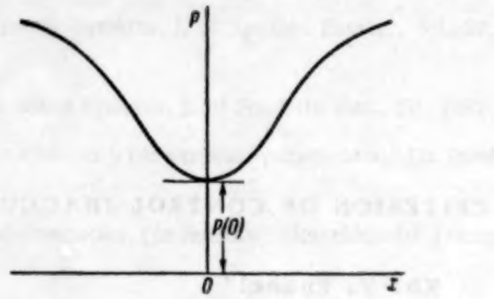


Fig. 2

relationship depends on the load on the boiler aggregate). In servo systems (where only the deviations of the controlled quantity from the supplied stimulus are measured, rather than the supplied stimulus itself), such arguments are frequently the supplied stimuli. Frequently, the variations of the extraneous arguments during use are random processes. Then, the importance P is a random function of x, and P = P(x) is called the mathematical expectation of this random function.

2. The Exact Investigation

We now derive exact formulae for the time average of the importance P and for the components of this average value which depend on the control process, from the assumptions that P is a function of the controlled quantity x and that this function is expressed in a Taylor series in the interval containing all the values of x that arise.

We now expand the importance P in a Taylor series in powers of x:

$$P = P(0) + P^{\mathrm{I}}(0)\,x + \frac{P^{\mathrm{II}}(0)}{2}x^2 + \frac{P^{\mathrm{III}}(0)}{6}x^3 + \frac{P^{\mathrm{IV}}(0)}{24}x^4 + \frac{P^{\mathrm{V}}(0)}{120}x^5 + \dots \tag{2}$$

Here, $P(0)$, $P^{\mathrm{I}}(0)$, $P^{\mathrm{II}}(0)$, $P^{\mathrm{III}}(0)$,... are the values of the importance and its derivatives with respect to x at the point x = 0.

We now integrate the series representing P between the limits of the time in use (0 to T) and divide by T:

$$\frac{1}{T}\int_0^T P\,dt = \frac{P(0)}{T}\int_0^T dt + \frac{P^{\mathrm{I}}(0)}{T}\int_0^T x\,dt + \frac{P^{\mathrm{II}}(0)}{2T}\int_0^T x^2\,dt + \frac{P^{\mathrm{III}}(0)}{6T}\int_0^T x^3\,dt + \frac{P^{\mathrm{IV}}(0)}{24T}\int_0^T x^4\,dt + \frac{P^{\mathrm{V}}(0)}{120\,T}\int_0^T x^5\,dt + \dots \tag{3}$$

We now consider this last expression is greater detail.

Its left member is the time average of the importance P:

$$\frac{1}{T}\int_0^T P\,dt = \overline{P}. \tag{4}$$

The right member is a series whose first term expresses the value of the importance at the point x = 0.

$$\frac{P(0)}{T}\int_0^T dt = P(0). \tag{5}$$

and the remaining terms of which contain various time averages: the time average of the deviation of the controlled quantity

$$\frac{1}{T}\int_0^T x dt = \bar{x}, \tag{6a}$$

the time average of the squared deviation of the controlled quantity

$$\frac{1}{T}\int_0^T x^2 dt = \overline{x^2}. \tag{6b}$$

etc.:

$$\frac{1}{T}\int_0^T x^3 dt = \overline{x^3}.$$

$$\frac{1}{T}\int_0^T x^4 dt = \overline{x^4}. \tag{6c}$$

$$\frac{1}{T}\int_0^T x^5 dt = \overline{x^5}. \tag{6d}$$

$$\cdots\cdots\cdots \tag{6e}$$

Thus, the time average of the importance is given in the form of the following infinite series:

$$\overline{P} = P(0) + P^{\mathrm{I}}(0)\,\bar{x} + \frac{P^{\mathrm{II}}(0)}{2}\,\overline{x^2} + \frac{P^{\mathrm{III}}(0)}{6}\,\overline{x^3} + \frac{P^{\mathrm{IV}}(0)}{24}\,\overline{x^4} + \frac{P^{\mathrm{V}}(0)}{120}\,\overline{x^5} + \dots \tag{7}$$

The component of the importance which depends on the control process

$$\bar{p} = \overline{P} - P(0) \tag{8}$$

takes the form of the following infinite series:

$$\bar{p} = P^{\mathrm{I}}(0)\,\bar{x} + \frac{P^{\mathrm{II}}(0)}{2}\,\overline{x^2} + \frac{P^{\mathrm{III}}(0)}{6}\,\overline{x^3} + \frac{P^{\mathrm{IV}}(0)}{24}\,\overline{x^4} + \frac{P^{\mathrm{V}}(0)}{120}\,\overline{x^5} + \dots \tag{9}$$

Expressions (7) and (9) are exact formulae.

3. The Approximate Theory

As shown by practice, the following three conditions (or at least some of them) can be considered to hold for almost all linear and nonlinear control systems.

Condition 1 (based on the fact that, at the point x = 0, the importance P has a minimum): for x = 0, the derivative of the importance with respect to the controlled quantity equals zero, $P^{\mathrm{I}}(0) = 0$.

Condition 2 (based on the fact that the deviations of the controlled quantity during use are, with equal

probability, on the increasing or decreasing side): the time average of the deviations of the controlled quantity and the time averages of the odd powers of the deviation of the controlled quantity are, in practice, equal to zero, $\overline{x} = 0$, $\overline{x^3} = 0$, $\overline{x^5} = 0,\ldots$

Condition 3 (based on the fact that the deviations of the controlled quantity are small): the time averages of the high powers of the deviations of the controlled quantity may be neglected, $\overline{x^3} \approx 0$, $\overline{x^4} \approx 0$, $\overline{x^5} \approx 0,\ldots$

These three conditions are important in obtaining approximate formulae from the exact formula (9) obtained above.

In formula (9), we may neglect the term $P^{I}(0)\overline{x}$ if conditions 1 or 2 hold: we may neglect the term $\frac{P^{III}(0)}{6}\overline{x^3}$ if conditions 2 or 3 hold; we may neglect the term $\frac{P^{IV}(0)}{24}\overline{x^4}$ if condition 3 holds; we may neglect the term $\frac{P^{V}(0)}{120}\overline{x^5}$ if conditions 2 or 3 hold, etc.

After these simplifications are made, the component of the time average of the importance which depends on the control procdss takes the following approximate form:

$$p \approx \frac{P^{II}(0)}{2}\overline{x^2}. \tag{10}$$

Since the factor $\frac{P^{II}(0)}{2}$ is a constant for a given controlled object and does not depend on the control process, the criterion for control inaccuracy turns out to be the time average of the squared deviation of the controlled quantity

$$\overline{x^2} = \frac{1}{T}\int_0^T x^2\,dt.$$

Instead of the criterion $\overline{x^2}$ one may use $\sqrt{\overline{x^2}}$ with the dimensionality of the controlled quantity, however, with this it is impossible to forget that $\sqrt{\overline{x^2}}$ imperfectly mirrors $\overline{p}$ (an increase of $\sqrt{\overline{x^2}}$ by factors of 2, 3, 4 etc. corresponds in actuality to a worsening of the "economics" of the process of control by factors of 4, 9, 16 etc.).

Methods and examples of computing $\overline{x^2}$ are given in [4-6]. Moreover, in certain works [6-10], there are presented such forms of the supplied and disturbing stimuli for which the integral of x^2 converges over an infinite interval of time. These works also provide methods of computing this integral, and examples; they may be useful in computing $\overline{x^2}$ when the stimuli are, for example, periodic or random functions of time.

SUMMARY

Criterion of control inaccuracy is theoretically well-grounded which makes it possible to objectively compare various automatic control systems.

LITERATURE CITED

[1] J. G. Truxal, Automatic Feedback Control System Synthesis, McGraw-Hill Book Co., 1955.

[2] S. D. Kuchkin, The Adjustment and Use of Combustion Process Regulators and ExKTI Systems for the Preparation of Pulverized Coal. [in Russian]. Gosénergoizdat (1954).

[3] V. L. Lossievskii, Automatic Control [in Russian] Published by the AN SSSR (1946).

[4] H.M. James, N.B. Nichols, and R.S. Phillips, Theory of Servomechanism [Mc Graw- Hill, 1947.

[5] V. V. Solodovnikov, Introduction to the Statistical Dynamics of Automatic Control Systems. [in Russian] Gostekhizdat (1952).

[6] A. A. Fel'dbaum, Electrical Systems for Automatic Control. [in Russian] Oborongiz (1957).

[7] E. G. Dudnikov, "A method for determining the optimum adjustment of industrial control systems from experimentally obtained dynamic characteristics," [in Russian]. In the collection "Automation of Industrial Processes" [in Russian]. Published by the AN SSSR (1955).

[8] A. A. Krasovskii, "Integral estimates and choice of parameters for automatic control systems," [in Russian]. In the book "Fundamentals of Automatic Control," [in Russian] edited by V. V. Solodovnikov, Mashgiz (1954).

[9] H. H. Rosenbrock, Integral of Error-Squared Criterion for Servomechanisms, Proc. IEE, pt. B, 102, No. 5, 1955.

[10] R. Voles, The "Least-Squares" Criterion Applied to Linear Servos. Electronic Eng. 26, No. 320, 1954.

Received December 15, 1958

THE CALCULATION OF PERIODIC MODES IN RELAY-TYPE AUTOMATIC CONTROL SYSTEMS

Yu. I. Alimov

(Sverdlovsk)

Exact methods [1, 2] are employed in the investigation of periodic modes in automatic control systems which consist of relay elements, linear links and inertialess functional transformers which do not appear in supplementary feedback loops and which are not shunted by them.

Examples of the use of this method are given. For relay-type systems of extremal control, the results of exact and of approximate calculation are compared.

1. In [1, 2] were developed an effective exact method for investigating periodic modes in relay-type automatic control systems, described by the following equations with respect to the representation:

$$
\begin{gathered}
D_i(p)\, Z_i(p) = K_i(p)\, Y(p) \quad (i = 1, \ldots, r). \\
Z(p) = \sum_{i=1}^{r} Z_i(p). \\
Y(p) = L\,\{y_0 + \Phi\,[z - z_0 + \tilde{f}(t)]\}, \\
\Phi(z) = \begin{cases} 0.5k_p\,[\operatorname{sign}(z - \varkappa_0) + \operatorname{sign}(z + \lambda\varkappa_0)] & \text{при} \quad \dot{z} > 0. \\ 0.5k_p\,[\operatorname{sign}(z + \varkappa_0) + \operatorname{sign}(z - \lambda\varkappa_0)] & \text{при} \quad \dot{z} < 0. \end{cases} \\
y_0,\ z_0,\ \varkappa_0,\ \lambda,\ k_p = \text{const}.
\end{gathered} \tag{1}
$$

where $K_i(p)/D_i(p)$ are the transfer functions of the linear links containing, in the general case, lags, distributed parameters, etc. and $\tilde{f}(t)$ is the periodic disturbance fed to the input of the relay link.

This method, based on simple frequency relationships and using the concept of the characteristic $J(\omega)$ of a relay system, was extended in [3] to the system in (1) when the nonlinear link is described by the equation

$$Y(p) = L\,\{az + Y_0 + \Phi\,[z - z_0 + \tilde{f}(t)]\} \quad (a = \text{const}).$$

We present below a method of computing the characteristic $J(\omega)$ for a relay system described by the set of equations

$$
\begin{aligned}
&Q_i(p)\, X_i(p) = R_i(p)\, e^{-p\tau_i}\, Y(p) && (i = 1, \ldots, r). \\
&U_j(p) = L\,\{F_j(x_1, \ldots, x_r)\}. \\
&Q_j(p)\, Z_j(p) = R_j(p)\, e^{p-\tau_j}\, U_j(p) && (j = r+1, \ldots, s).
\end{aligned} \tag{2}
$$

$$Y(p) = L\left\{\Phi\left[\sum_{j=1}^{s} z_j + \tilde{f}(t)\right]\right\}. \tag{2}$$

where $R_i(p)/Q_i(p)$ and $R_j(p)/Q_j(p)$ are linear fractional functions of p, in each of which the degree of the numerator is less than the degree of the denominator, and $F_j(x_1,\ldots,x_r) = -F_j(-x_1,\ldots,-x_r)$ are single-valued piece-wise continuous functions with piece-wise continuous partial derivatives with respect to any of their arguments.

By means of this characteristic, one may study the autooscillations ($\tilde{f}(t) \equiv 0$) of the system in (2), as well as its forced oscillations $[\tilde{f}(t) \not\equiv 0]$ in a forcing field analogously to the manner in which this was done in [1, 2] for the case of (1).

2. Initially, we consider the system described by the equations

$$\begin{gathered} Q_i(p)\,X_i(p) = R_i(p)\,e^{-p\tau_i}\,Y(p) \quad (i = 1,\ldots,r), \\ Z(p) = L\{F(x_1,\ldots,x_r)\}.\; Y(p) = L\{\Phi(z + \tilde{f}(t))\}. \\ \Phi(z) = \begin{cases} k_p \operatorname{sign}(z - \varkappa_0) & \text{for } \dot{z} > 0, \\ k_p \operatorname{sign}(z + \varkappa_0) & \text{for } \dot{z} < 0. \end{cases} \end{gathered} \tag{3}$$

where $\tau_i = 0$ $(i = 1,\ldots,r)$. We assume further that the function $F(x_1,\ldots,x_r)$ satisfies the requirements formulated above for the functions $F_j(x_1,\ldots,x_r)$.

If the equation $Q_i(p) = 0$ $(i = 1,\ldots r)$ has a simple root, then

$$\frac{R_i(p)}{Q_i(p)} = \sum_{\nu_i=1}^{n_i} \frac{c_{\nu_i}}{p - p_{\nu_i}}, \quad c_{\nu_i} = \frac{R_i(p_{\nu_i})}{Q_i'(p_{\nu_i})}.$$

where n_i is the degree of the polynomial $Q_i(p)$. Thus, by opening the system in (3) at the output of the relay element, and by denoting the output quantity of the latter by $v(t)$ for $\tilde{f}(t) \equiv 0$, we obtain

$$\begin{gathered} (p - p_{\nu_i})\,X_{\nu_i}(p) = c_{\nu_i} Y(p) \quad (i = 1,\ldots,r;\ \nu_i = 1,\ldots,n_i). \\ X_i(p) = \sum_{\nu_i=1}^{n} X_{\nu_i}(p),\; Z(p) = L\{F(x_1,\ldots,x_r)\}. \quad V(p) = L\{\Phi(z)\}. \end{gathered} \tag{4}$$

As is well known, the solution of the equation $(p - p_{\nu_i})\,X_{\nu_i}(p) = c_{\nu_i}\,Y(p)$ can be written in the form

$$x_{\nu_i}(t) = e^{p_{\nu_i} t}\left[x_{\nu_i}(0) + c_{\nu_i}\int_0^t y(\theta)\,e^{-p_{\nu_i}\theta}\,d\theta\right]. \tag{5}$$

Let $y(t) = \tilde{y}(t) = -\tilde{y}(t - \pi/\omega)$ be an infinite sequence of rectangular pulses with alternating polarity, height k_p and length π/ω, where $\tilde{y}(t) = k_p$ for $0 < t < \pi/\omega$. Then, the particular solution $x_{\nu_i}(t) = \tilde{x}_{\nu_i}(t)$, corresponding to the stable steady state of system (4), satisfies the condition $\tilde{x}_{\nu_i}(t) = -x_{\nu_i}(t + \pi/\omega)$. On the basis of (5) we may write

$$\tilde{x}_{\nu_i}(0) = -\tilde{x}_{\nu_i}\left(\frac{\pi}{\omega}\right) = \frac{k_p c_{\nu_i}}{p_{\nu_i}}\,\frac{1 - e^{p_{\nu_i}\pi/\omega}}{1 + e^{p_{\nu_i}\pi/\omega}} \quad \text{for } p_{\nu_i} \neq 0.$$

and, also,

$$-\frac{k_p c_{\nu_i}\pi}{2\omega} \qquad \text{for} \quad p_{\nu_i}=0. \tag{6}$$

By using (4) we find that

$$\tilde{\dot{x}}_{\nu_i}(-0) = -\tilde{\dot{x}}_{\nu_i}(\pi/\omega - 0) = \frac{2k_p c_{\nu_i} e^{p_{\nu_i}\pi/\omega}}{1+e^{p_{\nu_i}\pi/\omega}}. \tag{7}$$

To establish the proper switching conditions for a piece-wise smooth function, and even more so for a piece-wise continuous function, $F(x_1,\dots,x_r)$, is generally speaking, very complicated. The finding of these conditions can require [4] taking into account the small parameters of the system. Following [5], we shall assume in this paper that these conditions can be used in the form

$$\begin{gathered} \tilde{z}(-0) = F(\tilde{x}_1(-0),\dots,\tilde{x}_r(-0)) = \varkappa_0, \\ \tilde{\dot{z}}(-0) = \sum_{i=1}^{r}\left(\frac{\partial F}{\partial x_i}\right)_0 \tilde{\dot{x}}_i(-0) > 0, \end{gathered} \tag{8}$$

where

$$F[\tilde{x}_1(-0),\dots,\tilde{x}_r(-0)] = \lim_{t\to -0} F[\tilde{x}_1(t),\dots,\tilde{x}_r(t)].$$

$(\partial F/\partial x_i)_0 = \lim_{t\to -0} F'_{x_i}[\tilde{x}_1(t),\dots,\tilde{x}_r(t)]$ and $\tilde{x}_i(t)$ are determined in accordance with (5) and (6), and $x_i(-0)$ is determined in accordance with equation (7). Thus, the characteristic $J(\omega)$, by which it is possible to investigate the necessary conditions for the existence of simple symmetric autooscillations and forced periodic motions of the relay system of (3), is best determined [5] by the equation

$$J(\omega) = -\frac{1}{\omega}\tilde{\dot{z}}(-0) - j\tilde{z}(-0). \tag{9}$$

3. For the case $\tau_i \geq 0$ $(i = 1,\dots,r)$, the system in (4) is replaced by the system

$$\begin{gathered} (p-p_{\nu_i})X_{\nu_i}(p) = c_{\nu_i}e^{-p\tau_i}Y(p) \qquad (i=1,\dots,r;\ \nu_i = 1,\dots,n_i), \\ X_i(p) = \sum_{\nu_i=1}^{n_i} X_{\nu_i}(p),\ \ Y(p) = L\{\Phi[F(x_1,\dots,x_r)]\}, \end{gathered} \tag{10}$$

for which, with $k\pi/\omega + \tau_i \leq t < (k+1)\pi/\omega + \tau_i$ (k an integer),

$$\begin{gathered} \tilde{x}_{\nu_i}(t) = \begin{cases} (-1)^k k_p \dfrac{c_{\nu_i}}{p_{\nu_i}}\left\{\dfrac{2\exp[p_{\nu_i}(t-k\pi/\omega-\tau_i)]}{1+\exp(p_{\nu_i}\pi/\omega)} - 1\right\} & \text{for} \quad p_{\nu_i}\neq 0, \\ (-1)^k k_p c_{\nu_i}(-\pi/2\omega + t - k\pi/\omega - \tau_i) & \text{for} \quad p_{\nu_i} = 0. \end{cases} \\ \tilde{\dot{x}}_{\nu_i}(t) = (-1)^k k_p c_{\nu_i}\frac{2\exp[p_{\nu_i}(t-k\pi/\omega-\tau_i)]}{1+\exp(p_{\nu_i}\pi/\omega)}. \end{gathered} \tag{11}$$

If the frequency ω of the periodic stimulus $\tilde{y}(t)$ satisfies the inequalities $l\pi/\omega \le \tau_1 < (l+1)\pi/\omega$ (l an integer) then, as is easily seen, the functions $x_{\nu i}(t)$ are obtained from (11) for $0 \le t < (l+1)\pi/\omega - \tau_1 < \pi/\omega$ with

$$k = -(l+1), \tag{12}$$

and for $(l+1)\pi/\omega - \tau_1 \le t < \pi/\omega$ with

$$k = -l. \tag{13}$$

Conditions (8) for the proper moment and direction of switching, and the characteristic in (9) for the relay system in (10), are then found in accordance with (11) and (13). As is clear from (8), in computing the characteristic $J(\omega)$ for the system in (4), one may use the grapho-analytic methods [2, 6] for determining the characteristics of a relay system of the type given in (1).

4. If the system of (2) is opened at the output of the relay element, as was done above for the system in (3), then the open system thus obtained, when acted upon by the sequence of rectangular phase $y(t) = \tilde{y}(t) = -\tilde{y}(t+\pi/\omega)$ has a motion $\tilde{x}_i(t) = \sum_{\nu_i=1}^{n_i} \tilde{x}_{\nu i}(t)$ $(i = 1,\ldots,r)$ defined by the relationships (11), (12) and (13) such that the functions $\tilde{u}_j(t) = F_j[\tilde{x}_1(t),\ldots,\tilde{x}_r(t)]$ are known, in principle at least. Computation of the quantities $\tilde{z}(-0) = \sum_{j=1}^{s} \tilde{z}_j(-0)$, $\tilde{\tilde{z}}(-0) = \sum_{j=1}^{s} \tilde{\tilde{z}}_j(-0)$ then leads to the finding of the particular solutions $\tilde{z}_{\nu j}(t) = -\tilde{z}_{\nu j}(t+\pi/\omega)$ for the equation

$$(p - p_{\nu_j}) Z_{\nu_j}(p) = c_{\nu_j} e^{-p\tau_j} \tilde{u}_j(p) \quad (j = 1,\ldots, s;\ \ \nu_j = 1,\ldots, n_j).$$

which is computed in correspondence with (5). Particular interest inheres in the cases of piece-wise linear and step functions F_j, when the characteristic $J(\omega)$ is known to be computed in closed form.

5. By determining, by means of the characteristic in (9), the frequencies of the possible autooscillations, or the phase shift between the impressed oscillations and the internal periodic reactions in the system of (2), it is possible to investigate, based on the theorems in work [7], these modes for stability by variation equations. In case (3), the system, in a first approximation, is a linear pulsed system with constant coefficients, and all the reasoning with respect to stability of the periodic motions, given in [2] for systems of type (1), remains in force. If, in (2), the functions $F_j(x_1,\ldots,x_r)$ are approximated by portions of hyperplanes (in particular, when F_j is a piece-wise linear function of one argument) then the characteristic equation, $\det(\mathbf{U} - \lambda\mathbf{E}) = 0$, where $\mathbf{U}$ is the fundamental matrix of the variation matrix for system (2), may be obtained on the basis of the relationships given in [8]. As in the case of (1), the holding of the condition for the absence of additional switching inside a period [2, 5] is verified only by constructing the curve $\tilde{z}(t) + \tilde{f}(t)$, corresponding to the mode being studied. However, in many concrete problems, this turns out to be obviously superfluous.

In order to avoid unwieldy computations, we investigated above only the simple symmetric periodic motions of system (2) without any dead zones in the relay elements. With the use of the corresponding theorems from work [2], the reasoning given carries over to the cases of complex and asymmetric periodic modes and to the presence of dead zones. The computation of the characteristic of a relay system by the method given above is possible, in principle, for all systems consisting of relay elements, linear links and inertialess functional transformers, if these latter are not shufited by additional feedback loops (except inertia-less ones) and such loops, closed around the linear links, are not included in the circuit. With this, analogously to the particular case of (2), the expression for the characteristic $J(\omega)$ includes integrals of the form $\int_{t_1}^{t_2} \tilde{u}(t) e^{r_\nu t}\, dt$, where $\tilde{u}(t)$ is a

known function of time and r_ν is a complex number. Naturally, the computation is simplified if the transfer functions of the linear links of high order in systems (2) and (3), are approximated by sufficiently simple expressions, for example, by using the methods given in [9, 10].

We note finally that it is not advantageous in all concrete problems to make use of condition (8) or other conditions on switching by means of the relay system's characteristic.

6. We now give two examples of the application of the method presented for computing the periodic modes in some relay systems.

Example 1. Many methods of increasing the autooscillations frequency in relay systems may be considered [2] as the connection of accelerating elements in parallel in certain fundamental portions of the system in such a way that the input to the relay equals the sum of the outputs of the accelerating elements and the fundamental part of the system. With this, the characteristic of the relay system is the sum of the characteristics of the relay system without accelerating elements plus that of the system consisting only of the relay and the accelerating elements. In [2] there is computed the characteristic $J(\omega)$ for linear accelerating elements. In accordance with (5)-(9), this may also be carried out for accelerating elements which consist of series connections of linear portions and nonlinear transformers as occurs, for example, in optimal control systems [2, 11, 12]. Thus, for the nonlinear transformer $z_1 = a\,|x_1|\,x_1$ and linear portions with transfer functions

$$\frac{R_1(p)}{Q_1(p)} = \frac{X_1(p)}{Y(p)} = -\frac{k_1}{p} \text{ and } \frac{R_1(p)}{Q_1(p)} = -\frac{k_1}{T_1 p + 1}$$

($y(t)$ is the output quantity from the relay), the characteristics have the forms, respectively,

$$J_1(\omega) = -\frac{1}{\omega}\tilde{\dot{z}}_1(-0) - j\tilde{z}_1(-0) = k_p^2 k_1^2 a\left(-\frac{\pi}{2\omega^2} - j\frac{\pi^2}{4\omega^2}\right)$$

and

$$J_1(\omega) = k_p^2 k_1^2 a\left[2\,\mathrm{th}\,\frac{\pi}{2\omega T_1}\left(\mathrm{th}\,\frac{\pi}{2\omega T_1} - 1\right)\frac{1}{\omega T_1} - j\,\mathrm{th}^2\,\frac{\pi}{2\omega T_1}\right].$$

Example 2. We now compute the autooscillation frequency and the hunting loss in a relay system for extremal control with independent linear searches and an extremum indicator in the form of a derivative of the output quantity of the object. Following [13], we take the inertia of the latter into account by means of two first-order inertial links with time constants of T_1 and ξT_1. We shall assume that these links are connected to the input and output of an ideal object with the characteristic $u = -ax^2$, $a = \text{const}$. If we open the extremal control system (ECS) at the output of the controlling device, and denote the input and output quantities of the open system thus obtained by, respectively, $y(t)$ and $v(t)$, we are led to the following system of equations:

$$X(p) = W_1(p)\,Y(p).\quad U(p) = L\{-ax^2\}.\quad Z(p) = W_2(p)\,U(p).\quad V(p) = L\{\Phi(z)\}. \tag{14}$$

where $W_1(p) = k_1/(T_1 p + 1)\,p$, $W_2(p) = k_2 p/(\xi T_1 p + 1)$, k_1 and k_2 are constants and $\Phi(z)$ is the characteristic of the controlling device which, in the mode to be investigated of the closed ECS, is a generator of a simple periodic sequence of rectangular pulses of height $\pm k_p$.

If, in our notation, $y(t) = \tilde{y}(t) = -\tilde{y}(t + \pi/\omega)$ then, in accordance with (5) and (6), on the interval $0 < t < \pi/\omega$

$$\tilde{u}(t) = -a\tilde{x}(t)^2 = -ak_1^2 k_p^2\left[t - \left(\frac{\pi}{2\omega} + T_2\right) + \frac{2T_1\exp(-t/T_1)}{1 + \exp(-\pi/\omega T_1)}\right]^2,$$

$$\tilde{z}(t) = \exp\left(-\frac{t}{\xi T_1}\right)\left[\tilde{z}(0) + \frac{k_2}{\xi T_1}\int_0^t \frac{\partial\tilde{u}(t)}{\partial t}\exp\left(\frac{t}{\xi T_1}\right)dt\right]. \tag{15}$$

For the forced oscillation, $\tilde{z}(t) = \tilde{z}(t + \pi/\omega)$, we obtain the expressions

$$\tilde{z}(0) = \frac{2kT_1}{1-\exp\left(-\frac{\pi}{\omega\xi T_1}\right)}\Bigg\{\frac{\pi}{2\omega T_1} - 1 - \xi + \left[\frac{\pi}{2\omega T_1} + 1 + \xi + \frac{4}{(1-2\xi)\left[1+\exp\left(-\frac{\pi}{\omega\xi T_1}\right)\right]^2} - \right.$$

$$\left. - \frac{4-2\xi+\frac{\pi}{\omega T_1}(1-\xi)}{(1-\xi)^2\left[1+\exp\left(-\frac{\pi}{\omega T_1}\right)\right]}\right]\exp\left(-\frac{\pi}{\omega\xi T_1}\right) - \frac{4\exp\left(-\frac{2\pi}{\omega T_1}\right)}{(1-2\xi)\left[1+\exp\left(-\frac{\pi}{\omega T_1}\right)\right]^2} +$$

$$+ \frac{4-2\xi-\frac{\pi}{\omega T_1}(1-\xi)}{(1-\xi)^2\left[1+\exp\left(-\frac{\pi}{\omega T_1}\right)\right]}\exp\left(-\frac{\pi}{\omega T_1}\right)\Bigg\} = -\tilde{z}\left(\frac{\pi}{\omega}\right), \tag{16}$$

$$\dot{\tilde{z}}(0) = -\frac{1}{\xi T_1}\left[\tilde{z}(0) - k_2\tilde{u}(0)\right], \quad k = -k_p^2 k_1^2 a k_2. \tag{17}$$

The expression for $\tilde{z}(0)$ for $\xi = 0.5$ ($\xi = 1.0$) is obtained directly from (16) by going to the limit $\xi \to 0.5$ ($\xi \to 1.0$).

The frequency of the possible autooscillations is determined from the conditions for the proper moment and direction of switching, which are conveniently written in the form

$$\frac{\tilde{z}(0)}{kT_1} = \varkappa_0 = \frac{z_0}{kT_1}, \tag{18}$$

$$\frac{\xi\dot{\tilde{z}}(0)}{k} = 2\,\mathrm{th}\,\frac{\pi}{2\omega T_1}\left(\frac{\pi}{2\omega T_1} - \mathrm{th}\,\frac{\pi}{2\omega T_1}\right) - \varkappa_0 > 0, \tag{19}$$

where z_0 is the threshold level of the input coordinate of the controlling device of the relay ECS ($z_0 < 0$ for $a > 0$). The solution of equation (18) can be obtained by means of the curves

$$\varkappa(\omega T_1, \xi) = \frac{\tilde{z}(0)}{kT_1}. \tag{20}$$

For $\xi > 0$, condition (19) always holds.

The hunting loss

$$\Delta_z = \frac{k_2\omega}{\pi}\left|\int_0^{\pi/\omega}\frac{\partial\tilde{u}(t)}{\partial t}\,dt\right|$$

as a function of the scanning frequency ω is determined, according to (15), by the expression

$$\Delta = \frac{\Delta_z}{kT_1^2} = \frac{\pi^2}{12\omega^2T_1^2} - 1 + \frac{2\omega T_1\left[1-\exp\left(-\frac{2\pi}{\omega T_1}\right)\right]}{\pi\left[1-\exp\left(-\frac{\pi}{\omega T_1}\right)\right]}.$$

The approximate formula

$$\Delta = \frac{8}{\pi^2\omega^2 T_1^2 (1 + \omega^2 T_1^2)^2}, \tag{21}$$

is obtained in [13] by the method of harmonic balance and, in the range $0 < \omega T_1 < 1.5$, gives results very close to the exact ones.

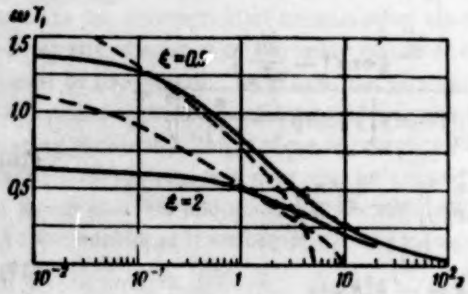


Fig. 1

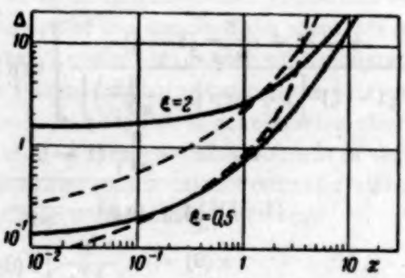


Fig. 2

By obtaining graphs constructed on the basis of expressions (20) and (21), one may easily find the relationship $\Delta = \Delta(\kappa, \xi)$ giving the hunting loss as a function of the threshold value of the controlling device's input quantity.

Figures 1 and 2 show, respectively, the curves constructed in accordance with equation (20) and the curves $\Delta = \Delta(\kappa, \xi)$ for $\xi = 0.5$ and $\xi = 2.0$, calculated in accordance with (16) (the solid lines) and by the approximate formula from [13]

$$\kappa = \frac{32\,(1 + \xi - \xi\omega^2 T_1^2)}{\pi^2 (1 + \omega^2 T_1^2)^2 (1 + 4\xi^2\omega^2 T_1)^2} \tag{22}$$

(the dotted lines).

It should be mentioned here that in the construction of the graphs of (22) for the range $\omega T_1 < 0.1$ to 0.2, inaccuracies were admitted in [13].

It is clear from Figs. 1 and 2 that the accuracy of the approximate formula from [13] increases somewhat with decreasing ξ and falls both for small and for large values of ωT_1, remaining comparatively high only in a rather narrow region of the scanning frequency, as a function of ξ. As a whole, the method of harmonic balance as applied to the ECS considered here gives low computational accuracy. This is affected by the fact that, in an ECS with an extremum indicator in the form of a derivative, the higher harmonics of the system's coordinate frequently have a significant value. It should be expected that in an ECS where the extremum indicator uses the deviation of the object's output quantity from an extremal value, and even more so when the integral over time of these deviations is used [14], the method of harmonic balance provides better agreement with the

SUMMARY

Accurate methods [1, 2] of analyzing periodic operation conditions are applied to automatic control systems which consist of relay part, linear units and of inertialess functional generators when latter are not included in additional feedbacks.

Examples of using the described method are given. Accurate and approximate calculation results are compared for an extremum control relay system.

Received December 2, 1958

LITERATURE CITED

[1] Ya. Z. Tsypkin, Transient Responses and Steady-State Processes in Pulsed Circuits. [in Russian]. Gosénergoizdat (1951).

[2] Ya. Z. Tsypkin, Theory of Relay Systems of Automatic Control. [in Russian]. Gostekhizdat (1955).

[3] M. A. Aizerman and F. R. Gantmakher, "One one class of dynamic problems, leading to the theory of relay systems," [in Russian]. Prikl. Matem. i Mekh. 19, 2 (1955).

[4] M. A. Aizerman and F. R. Gantmakher, "On certain switching peculiarities in nonlinear automatic control systems with a piece-wise smooth characteristic for the nonlinear element," [in Russian]. Automation and Remote Control (USSR) 18, 11 (1957).

[5] N. A. Korolev, "On periodic modes in relay systems with internal feedback loops," [in Russian]. Automation and Remote Control (USSR) 17, 11 (1956).

[6] L. P. Kuz'min, "A grapho-analytic method for determining the characteristic of a relay system," [in Russian]. Automation and Remote Control (USSR) 19, 4 (1958).

[7] M. A. Aizerman and F. R. Gantmakher, "Stability of a first approximation periodic solution of a system of differential equations with discontinuous right members," [in Russian]. Prikl. Matem. i Mekh. 21, 5 (1957).

[8] M. A. Aizerman and F. R. Gantmakher, "On the stability of the periodic modes in nonlinear systems with piece-wise linear characteristics," [in Russian]. Automation and Remote Control (USSR) 19, 5 (1958).

[9] M. P. Simiyu, "Determining the coefficients of the transfer functions of the linearized links in auto-control systems," [in Russian]. Automation and Remote Control (USSR) 18, 6 (1957).

[10] A. A. Kardashev and L. V. Kornyushin, Determination of system parameters by experimental (given) frequency characteristics," [in Russian]. Automation and Remote Control (USSR) 19, 4 (1958).

[11] A. A. Fel'dbaum, "Optimal processes in automatic control systems," [in Russian]. Automation and Remote Control (USSR) 14, 6 (1953).

[12] A. A. Fel'dbaum, "On synthesizing optimal systems by means of phase space," [in Russian]. Automation and Remote Control (USSR) 16, 2 (1955).

[13] I. S. Morosanov, "Methods of extremal control," [in Russian]. Automation and Remote Control (USSR) 18, 11 (1957).

[14] Hsüeh-Sen Tsien, Engineering Cybernetics. [Russian translation]. IL (1956).

THE USE OF NONLINEAR CORRECTING DEVICES OF THE "KEY" TYPE FOR IMPROVING THE QUALITY OF SECOND-ORDER AUTOMATIC CONTROL SYSTEMS

S. V. Emel'yanov
(Moscow)

The paper deals with the questions of stabilization and improvement of regulation quality of linear second-order control systems by means of nonlinear correcting devices of the "key" type.

In [1] are presented the principles of constructing a nonlinear corrector, based on the nonlinear transformation of functions of several variables by means of uniform "key"-type devices. It was shown that, by means of such devices, the most diverse control laws can be realized.

The present paper is devoted to the questions of using a nonlinear corrector for stabilizing, and improving the regulation of, linear automatic control systems. We shall consider closed dynamic systems described by second-order equations. For the determination of the tuning of the correcting device we have chosen the phase plane method [2]. In particular, we use the concept of the many-sheeted phase plane which was used in the investigation of second-order systems by V. V. Petrov, and G. M. Ulanov [3-5], and also in the works of V. V. Kazakevich.

For the class of systems considered we select the minimum number of elementary "key"-type (Ψ-cell) devices, namely, one Ψ-cell.

Using the phase plane method, we carry out a qualitative investigation of such nonlinear dynamic systems. We shall determine all the possible forms of motion after any initial stimulus. Based on this, we shall find the formula for the optimum tuning of the corrector, determing the relationship between the parameters of the correcting device and the system parameters for which the control system, with type of correcting device considered, will possess the best transient response quanlity. The problem posed here leads to the study of the phase plane of the system being considered and to the structure of its partioning into different forms as one varies the coefficients of the equation of motion and the initial position of the representative point.

The Equations of Motion

The equations of motion of the systems considered may be written in the following way:

the object equation

$$T_a x + \rho x = -\mu + f(t), \tag{1}$$

the equation for the driving device

$$x - g(t) = \varphi, \tag{2}$$

the equation for the error signal amplifier

$$\eta = \frac{1}{\delta}\varphi, \tag{3}$$

the equation for the "key"-type ($\overline{\Psi}$-cell) correcting device

$$\chi = \Psi(\varphi, \dot{\varphi})\,\varphi, \tag{4}$$

the adder equation

$$\xi = \eta - \chi, \tag{5}$$

the servomotor equation

$$T_s\dot{\mu} = \xi. \tag{6}$$

Here, x is the relative deviation of the controlled coordinate, φ is the relative derivation of the error signal, χ is the relative deviation of the output coordinate of the correcting device, ξ is the relative variation of the adder's input coordinate, μ is the relative deviation of the controlling organ, T_a is the acceleration time, ρ is a coefficient which characterizes the object's self-smoothing, T_s is the regulator's integration constant, δ is the amplifier's irregularity difficient, $\Psi(\varphi, \dot{\varphi})$ is the nonlinear transmission factor of the correcting device, f(t) is the external disturbance and g(t) is the driving stimulus.

The block schematic of the system is shown in Fig. 1.

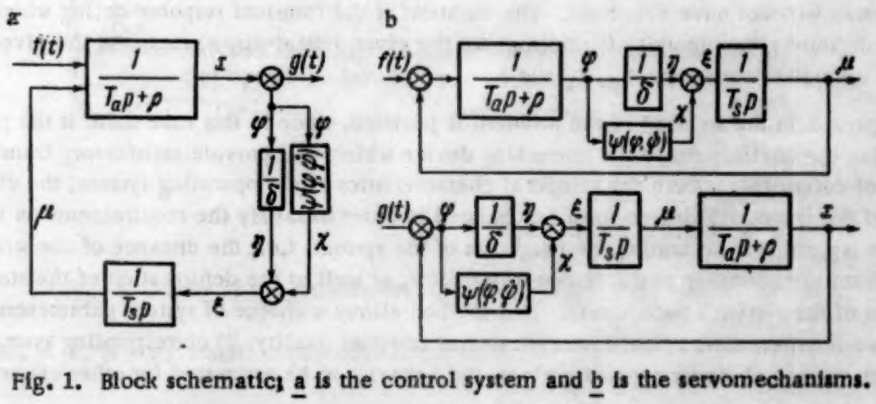


Fig. 1. Block schematic; a is the control system and b is the servomechanisms.

The value of the nonlinear transmission factor $\Psi(\varphi, \dot{\varphi})$ is expressed in terms of the constant coefficients of the nonlinear corrector, k_i, k_j, T_j (Fig. 2) in the following way:

$$\Psi(\varphi, \dot{\varphi}) = \begin{cases} k_i & \text{for } (T_j\dot{\varphi} + k_j\varphi)\,\varphi \leqslant 0 \\ 0 & \text{for } (T_j\dot{\varphi} + k_j\varphi)\,\varphi > 0. \end{cases}$$

Investigation of the Dynamics of a Second-Order Automatic Control System with a Nonlinear Correction of the Form $\Psi(\varphi, \dot{\varphi})\,\varphi(t)$

1. An Object Without Self-smoothing ($\rho = 0$).

We now consider in detail the case when the controlled object possesses no self-smoothing ($\rho = 0$). With

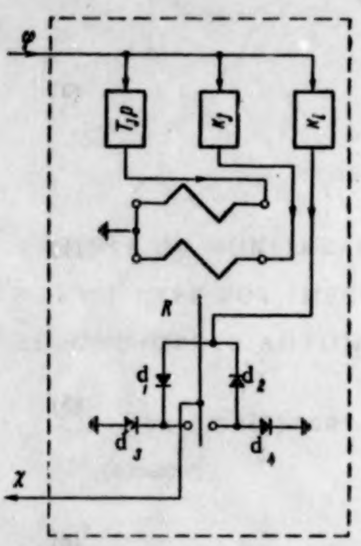


Fig. 2. Circuit of the nonlinear "key"-type correcting device. R is the type RP-4 or RP-5 relay, D_1, D_2, D_3 and D_4 are semiconductors of type DGTs-24, DGTs-27 or D-2-E and $p = d/dt$.

this simple example we shall try to show the effectiveness of the correcting method suggested, and also to elucidate certain peculiar features inherent in such nonlinear systems.

By eliminating the variables x, μ, η, χ, ξ, from equations (1)-(6), we obtain the nonlinear differential equation

$$\ddot{\varphi} + (\omega_0 - \bar{\beta})\varphi = 0, \tag{8}$$

where

$$\omega_0 = \frac{1}{T_a T_s \delta}, \quad \bar{\beta} = \frac{\Psi(\varphi, \dot{\varphi})}{T_a T_s}.$$

This equation is valid for any initial disturbance and also for $f(t) = C$, $g(t) = C$, where C is a constant.

We consider the dynamics of the system, described by nonlinear equation (8), in the following order.

a) We partition the space of the parameters ω_0, β, J_j ($\rho = k_i/T_a T_s$, $J_j = k_j/T_j$) into a series of representative cross sections which will be the planes of the parameters ω_0, β, for constant values of the coefficient J_j, and we shall determine the character of the motion of the representative point on the phase planes in the regions into which the corresponding planes of the parameters ω_0 and β may be partitioned, as well as on the boundaries between these regions.

b) We shall choose the values of the coefficients of the correcting device (Ψ-cell), k_i, k_j and T_j, for which the transient response will not have overshoot. The duration of the transient response during which the error is made less than a definite given quantity is minimal for the given initial disturbance and the given coefficients of the object and controller equations, T_a, T_s and δ.

Such an approach to the solution of the problem is justified, since in this case there is the possibility, not only of determining the coefficients of the correcting device which will provide satisfactory transient response quality, but also of determining, from the temporal characteristics of the operating system, the direction in which the coefficients of the correcting device must be changed in order to satisfy the requirements on transient response quality. It is possible to determine the roughness of the system, i.e., the distance of the working point on the ω_0, β plane from the boundary of the region of stability, as well as the deformation of the stability boundaries as a function of the system's parameters. This method allows a choice of system parameters which simultaneously meet two requirements: 1) satisfactory transient response quality; 2) corresponding system roughness. It also allows, with the use of linear approximations, the behavior to be estimated for other classes of nonlinear systems.

We shall consider three representative cross-sections of the space of the parameters, ω_0, ρ, J_j namely, the planes of the parameters ω_0 and β for $J_j = 0$, $\infty > J_j > 0$ and $J_j = \infty$. The plane of the parameters ω_0, β for $J_j = 0$ corresponds to a zero value for the coefficient k_j. In other words, the relay winding (Fig. 2) has impressed upon it a signal which is proportional only to the derivative of the controlled coordinate.

The Cross Section for $J_j = 0$ ($k_j = 0$, $T_j \neq 0$)

The motion of the representative point will occur on a double-sheeted phase plane. The boundaries of the sheets are defined by the equation $\varphi\dot{\varphi} = 0$, i.e., in other words, $\varphi = 0$ and $\dot{\varphi} = 0$.

Sheet I is the part of the phase plane satisfying the condition $\varphi\dot{\varphi} < 0$ and sheet II is the portion of the phase plane satisfying the condition $\varphi\dot{\varphi} > 0$. The plane of parameters ω_0 and β (Fig. 3) falls into a number of regions which are distinguished by the different forms of motion of the representative point on the phase plane.

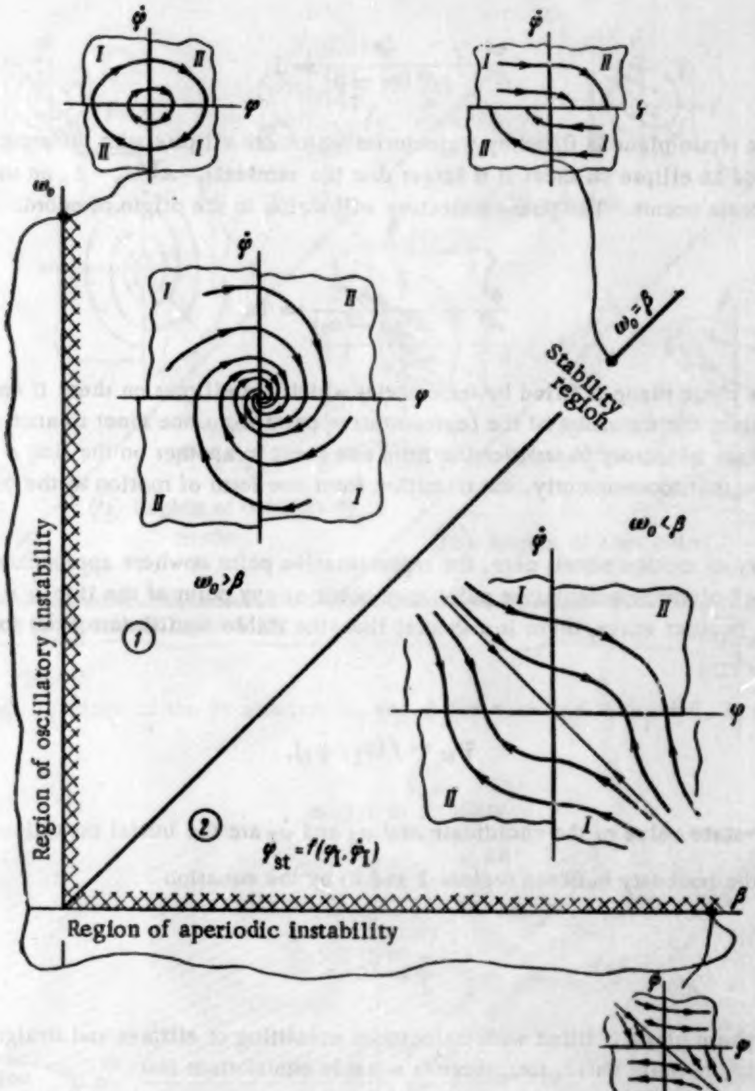


Fig. 3. Plane of the parameters ω_0 and β for $I_j = 0$ ($k_j = 0$, $T_j \neq 0$).

Case 1. $\omega_0 > 0$, $\beta = 0$. Equation (8) reduces the equation

$$\ddot{\varphi} + \omega_0 \varphi = 0. \tag{9}$$

The phase plane will be filled with concentric ellipses. The system will lie on the boundary of stability.

Case 2. $\omega_0 > 0$, $\beta > 0$. The motion of the representative point on sheet II is defined by the equation

$$\frac{\varphi^2}{A^2} + \frac{\dot{\varphi}^2}{A^2 \omega_0} = 1. \tag{10}$$

where A is an arbitrary constant, defined by [the initial conditions since $\Psi(\varphi, \dot{\varphi}) = 0$ for $\varphi\dot{\varphi} > 0$].

On sheet I, the motion of the representative point will be defined by the following equations.

a) For $\omega_0 > \rho$ (region 1), by the equation

$$\frac{\varphi^2}{A^2} + \frac{\dot{\varphi}^2}{A^2(\omega_0 - \beta)} = 1. \tag{11}$$

In this region the phase plane is filled by trajectories which are ellipses with different semiaxes. Since the semiaxis, $A\sqrt{\omega_0}$, of an ellipse on sheet II is larger that the semiaxis, $A\sqrt{\omega_0 - \beta}$, on sheet I, the case of a damped oscillatory process occurs. The phase trajectory will shrink to the origin of coordinates.

$$\frac{\varphi^2}{A^2} - \frac{\dot{\varphi}^2}{A^2(\beta - \omega_0)} = 1. \tag{12}$$

In this region the phase plane is filled by trajectories which are ellipses on sheet II and hyperbolae on sheet I. As in the previous case, the transition of the representative point from one sheet to another occurs on the lines $\varphi = 0$ and $\dot{\varphi} = 0$, the phase trajectory in transferring from one sheet to another on the line $\dot{\varphi} = 0$ being directed from one to the other so that, consequently, the transition from one form of motion to the other is a continuous process.

Since the velocity of motion equals zero, the representative point nowhere approaches the origin of coordinates. Such a "sticking" of the representative point may occur at any point of the line $\dot{\varphi} = 0$ as a function of the initial conditions. In other words, there is a special line (the stable equilibrium line) for which the following equation is valid

$$\varphi_{st} = f(\varphi_I, \dot{\varphi}_I), \tag{13}$$

where φ_{st} is the steady-state value of the coordinate and φ_I and $\dot{\varphi}_I$ are the initial conditions.

c) For $\omega_0 = \beta$ (the boundary between regions 1 and 2) by the equation

$$\ddot{\varphi} = 0. \tag{14}$$

In this case, the phase plane is filled with trajectories consisting ot ellipses and straight lines parallel to the φ axis. Equation (13) remains valid, i.e., there is a stable equilibrium line.

Case 3. $\omega_0 = 0$, $\rho > 0$. The behavior of the system under these conditions has certain peculiarities. The equation of the phase trajectories on sheet II degenerates to equation (14). The motion of the representative point on sheet I is defined by equation (12).

The stability of such a system depends on the initial conditions. A stable motion corresponds to each set of initial conditions, φ_I and $\dot{\varphi}_I$, which lies in the plane of sheet I between the line $\dot{\varphi} = 0$, including the line itself, and the null phase trajectory hyperbola. Unstable motions correspond to each set of initial conditions, φ_I $\dot{\varphi}_I$, lying in the plane of sheet II and also in that portion of the plane of sheet I lying between the line $\varphi = 0$, including the line itself, and the null phase trajectory hyperbola. Equation (13) is valid for those initial conditions to which stable motions correspond.

It is not difficult to show that the cases $\omega_0 < 0$, $\beta \gtrless 0$ and $\omega_0 > 0$, $\beta < 0$ correspond to unstable system motions, where the case $\omega_0 > 0$, $\beta < 0$ corresponds to the region of oscillatory instability and the case $\omega_0 < 0$, $\beta \gtrless 0$ corresponds to the region of aperiodic instability.

The Cross Section for $\infty > J_j > 0$ ($k_j \neq 0$, $T_j \neq 0$)

The motion of the representative point in this case will occur on a two-sheeted phase plane. The boundaries of the sheets are defined by the equations $\varphi = 0$ and $\dot{\varphi} = -J_j\varphi$. Sheet I is the portion of the phase plane which satisfies the condition $(T_j\dot{\varphi} + k_j\varphi)\varphi > 0)$.

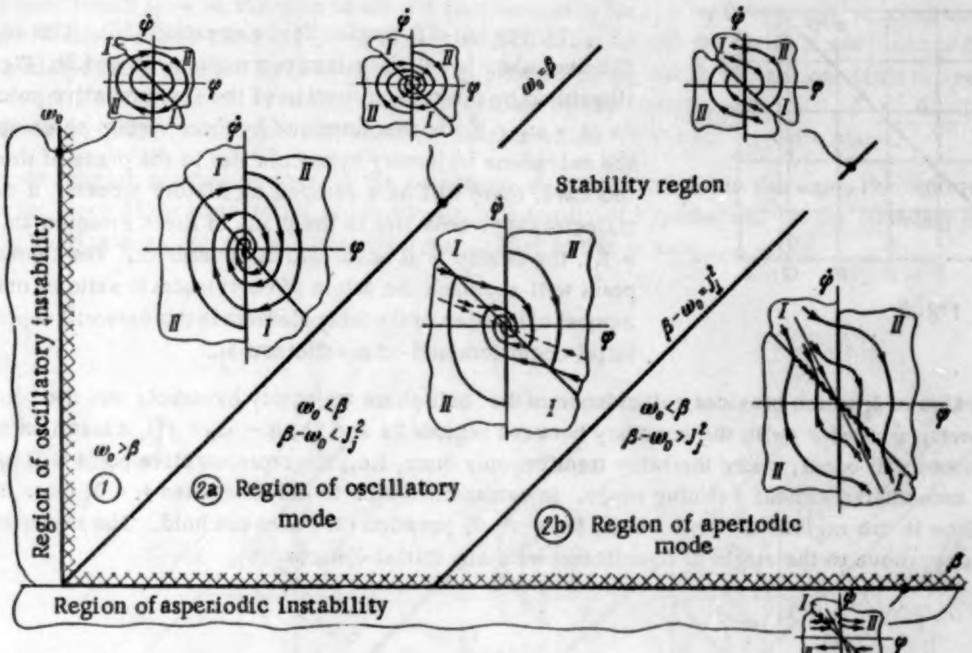


Fig. 4. Plane of the parameters ω_0 and β for $\infty > J_j > 0$ ($k_j \neq 0$, $T_j \neq 0$).

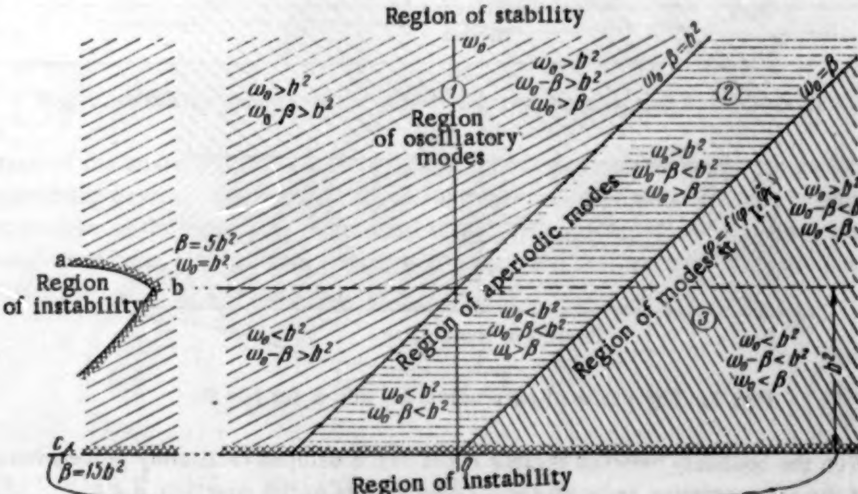


Fig. 5. The parameter plane for $J_j = 0$.

We now consider a cross section of the space of the parameters ω_0, β and J_j by a plane with the value of J_j lying between the limits $\infty > J_j > 0$, and note the differences in system behavior in the corresponding regions of the plane of the parameters ω_0 and ρ with respect to the section for $J_j = 0$ (Fig. 4).

Case 1. $\omega_0 > 0$, $\beta = 0$. Just as in the previous case, the phase plane will be filled by concentric ellipses. The system will be on the boundary of stability for any value of the coefficient J_j.

Case 2. $\omega_0 > 0$, $\beta > 0$. The motion of the representative point on sheet I is defined by equation (10).

On sheet I, the motion of the representative point is determined in the following way.

a) For $\omega_0 > \beta$ (region 1), by equation (11). As in the case when $J_j = 0$, a damped oscillatory process

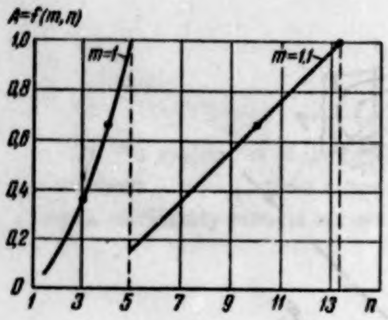


Fig. 6

will occur, the tendency of the system to oscillate increasing (for constant values of ω_0 and β) with increasing J_j.

b) For $\omega_0 < \beta$ (region 2), by equation (12). This region unlike the case when $J_j = 0$, falls into two regions, 2a and 2b (Fig. 4), distinguished by the type of motion of the representative point. Region 2a ($\beta - \omega_0 < J_j^2$) is characterized by those system parameters for which the null-phase trajectory hyperbola lies in the plane of sheet II. In this case, there will be a damped oscillatory process. If the null-phase trajectory hyperbola lies in the plane of sheet I (region 2b, $\beta - \omega_0 > J_j^2$), the process will occur without overshoot. The representative point will approach the origin of coordinates in a sliding mode. the number of transfers of the relay element in this case will depend on the intial conditions and on coefficients J_j.

For a value of J_j which provides coincidence of the null-phase trajectory hyperbola and the boundary between the sheets, $\dot{\varphi} = -J_j\varphi$ (with the boundary between regions 2a and 2b, $\beta - \omega_0 = J_j^2$), a transient response without overshoot will occur, where the relay transfers only once, i.e., the representative point will approach the origin of coordinates without a sliding mode. In contradistinction to the case when $J_j = 0$, there is no stable equilibrium line in this region. In other words, for $J_j \neq 0$, equation (13) does not hold. The representative point will always move to the origin of coordinates with any initial disturbance.

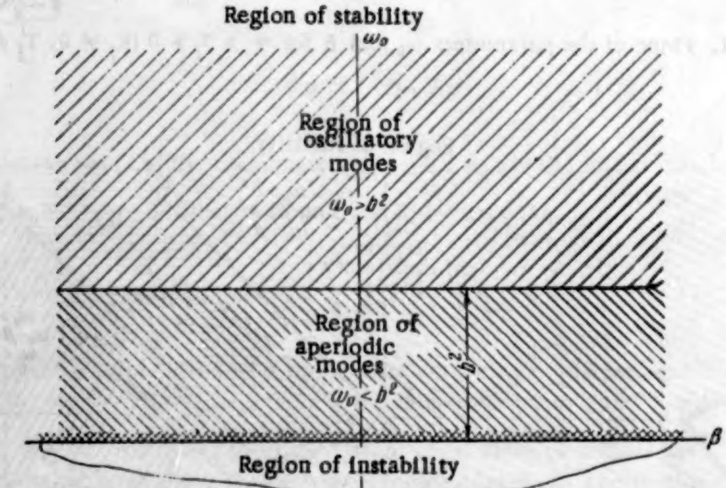


Fig. 7. The plane of parameters ω_0 and β for $J_j = \infty$.

c) For $\omega_0 = \beta$ (on the boundary between regions 1 and 2a), a damped oscillatory process occurs. The equation of motion of the representative point on sheet I degenerates to the equation $\ddot{\varphi} = 0$.

Case 3. $\omega_0 = 0$, $\beta > 0$. The behavior of the system under these conditions possesses the following special features. If the null-phase trajectory hyperbola, in mirroring the motion of the representative point on sheet I, lies in the plane of sheet I or on the boundary between the sheets, $\dot{\varphi} = -J_j\varphi$, then system stability will depend on the initial conditions. A stable motion corresponds to each set of initial conditions, φ_I and $\dot{\varphi}_I$, which lies in the plane of sheet I between the null phase trajectory hyperbola and the boundary $\dot{\varphi} = -J_j\varphi$, and also to all sets of initial conditions lying in the plane of sheet II between the line $\dot{\varphi} = 0$ and the boundary $\dot{\varphi} = -J_j\varphi$. To all initial conditions which lie outside of these regions, an unstable motion corresponds. If the null-phase trajectory hyperbola does not lie in the plane of sheet I then, for any initial conditions, the motion will be unstable.

It is easily shown that the cases $\omega_0 < 0$, $\beta \gtrless 0$ and $\omega_0 > 0$, $\beta < 0$ correspond to unstable system motions for any value of J_j, where the case $\omega_0 > 0$, $\beta < 0$ corresponds to the region of oscillatory instability and the case $\omega_0 < 0$, $\beta \gtrless 0$ corresponds to the region of aperiodic instability.

The Cross-Section for $J_j = \infty$ ($k_j \neq 0$, $T_j = 0$)

As the coefficient $J_j \to \infty$, the area of sheet I decreases and, for $J_j = \infty$, it becomes equal to zero, i.e., the boundaries dividing sheets I and II ($\varphi = 0$ and $\dot{\varphi} = -J_j\varphi$) coincide and the entire phase plane is filled with trajectories corresponding to sheet II. The plane of the parameters ω_0 and β is partitioned into two regions by the form of the phase trajectories. The first region is defined by the conditions $\omega_0 > 0$, $\beta \gtrless 0$. The motion of the representative point in this case will occur by ellipses which fill the entire phase plane.

The second region is characterized by the conditions $\omega_0 < 0$, $\beta \gtrless 0$. In this case, the representative point will move along phase trajectories which constitute a family of hyperbolae. On the boundary between these regions, the phase trajectories are straight lines parallel to the φ axis.

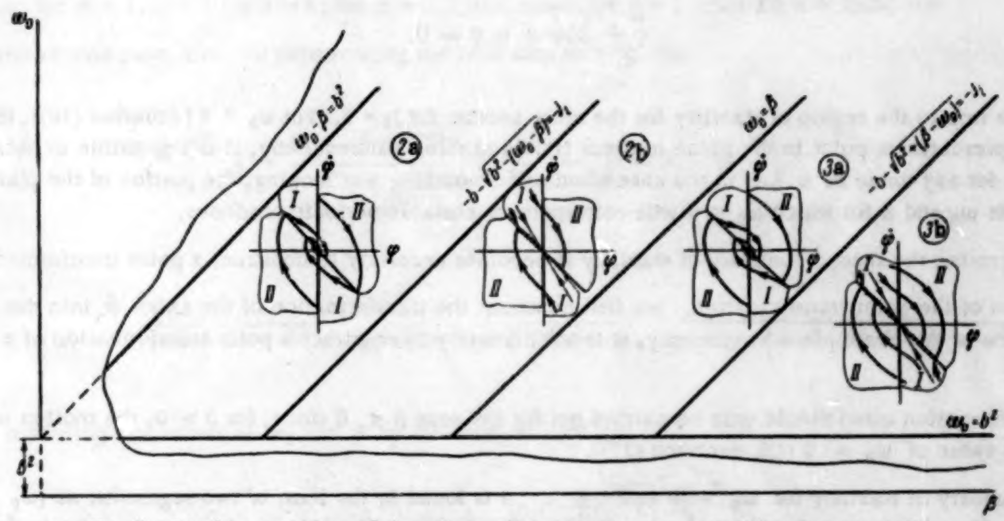


Fig. 8. Working portion of the plane of parameters ω_0 and β for $\infty > J_j > 0$.

On the basis of the analysis just made, one can formulate certain recommendations with respect to the tuning of the correcting device. The formula for the optimal tuning of the corrector, defining the relationship between the parameters of the correcting device and those of the system, for which the correcting device, of the structure considered here, for a control system will possess the best transient response qualities, is determined from the condition $\beta - \omega_0 = J_j^2$ and has the form

$$k_1 = \frac{1}{\delta} + T_a T_s \left(\frac{k_j}{T_j}\right)^2. \tag{15}$$

2. An Object with Positive Self-Smoothing

We now consider the case, a very frequent one in practice, when an object with positive self-smoothing ($\rho > 0$) is controlled by variable-speed servomotors without rigid feedback loops. In this case, after eliminating the variables x, μ, η, χ, and ξ from equations (1)-(6), we obtain a nonlinear differential equation of the form

$$\ddot{\varphi} + 2b\dot{\varphi} + (\omega_0 - \bar{\beta})\varphi = 0. \tag{16}$$

where

$$2b = \frac{\rho}{T_a}.$$

The order of considering the dynamics (free oscillations) of the system we shall keep the same as in the previous case for ease of comparison.

The Cross-Section for $J_j = 0$ ($k_j = 0$, $T_j \neq 0$)

The motion of the representative point in this case will also occur on a two-sheeted phase plane. The boundaries of the sheets are defined by the equation $\varphi\dot{\varphi} = 0$, i.e., by the two equations $\varphi = 0$ and $\dot{\varphi} = 0$. On sheet I ($\varphi\dot{\varphi} < 0$), the equations of the phase trajectories are obtained by eliminating time from the equation

$$\ddot{\varphi} + 2b\dot{\varphi} + (\omega_0 - \beta)\varphi = 0. \tag{17}$$

On sheet II, by eliminating time from the equation

$$\ddot{\varphi} + 2b\dot{\varphi} + \omega_0\varphi = 0. \tag{18}$$

We now isolate the region of stability for the cross section for $J_j = 0$. For $\omega_0 < 0$ [equation (18)], the motion of the representative point in the plane of sheet II be unstable. Consequently, it is impossible to obtain a stable system for any value of β. As in the case when self-smoothing was lacking, the portion of the plane of the parameters ω_0 and ρ for which $\omega_0 < 0$ will correspond to unstable periodic processes.

To determine the second boundary of stability it becomes necessary to construct a point transformation.

Equation of the point transformation. We first construct the transformation of the axis $+\dot{\varphi}$ into the axis $-\dot{\varphi}$. (By virtue of the phase plane's symmetry, it is not necessary to construct a point transformation of a line into itself.)

The construction cited should only be carried out for the case $\beta < 0$ since, for $\beta > 0$, the motion will be stable for any value of $\omega_0 > 0$ [Cf. equation (17)].

The boundary of stability for $\omega_0 > 0$ and $\beta < 0$ is found in the form of two segments: ab ($\omega_0 > 0$, $\omega_0 \geq b^2$, $\beta < 0$) and bc ($\omega_0 > 0$, $\omega_0 \leq b^2$, $\beta < 0$) (Fig. 5). The motion of the representative point in the plane of sheet II is described, for these cases, by different equations, which are obtained by eliminating time from equation (18) for different values of the coefficients ω_0 and b. The motion of the representative point on the plane of sheet I is described by equations obtained in the same manner, but from equation (17).

The transformation mentioned is found in the form

$$\dot{\varphi}_k^2 = A\dot{\varphi}_I^2. \tag{19}$$

where $\dot{\varphi}_I$ is the initial value of velocity on the half-line $+\dot{\varphi}$, $\dot{\varphi}_k$ is a finite value of velocity on the half-line $-\dot{\varphi}$ and A is some constant coefficient which is expressed in terms of the system's parameters.

We note that $A < 1$ if the system is stable and $A > 1$ if the system is unstable; if $A = 1$, the system will lie on the boundary of stability.

For the first case, i.e., for determining the boundary ab (Fig. 5):

$$A = \frac{b^2 + \omega_1'^2}{b^2 + \omega_1^2} \exp\frac{2b}{\omega_1}\left[\operatorname{arc\,tg}\frac{b}{\omega_1} - \frac{\pi}{2}\right] \exp\frac{2b}{\omega_1'}\left[-\frac{\pi}{2} - \operatorname{arc\,tg}\frac{b}{\omega_1'}\right]. \tag{20}$$

where

$$\omega_1' = \sqrt{(\omega_0 + \beta) - b^2}, \quad \omega_1 = \sqrt{\omega_0 - b^2}.$$

We now express ω_0 and β in terms of b^2. Let $\omega_0 = mb^2$ and $\beta = nb^2$, where m and n are arbitrary numbers. By substituting these values in (20) we get

$$A = \frac{m+n}{m} \exp \frac{2}{\sqrt{m-1}} \left[\operatorname{arc\,tg} \frac{1}{\sqrt{m-1}} - \frac{\pi}{2} \right] \times$$

$$\times \exp \frac{2}{\sqrt{(m+n)-1}} \left[-\frac{\pi}{2} - \operatorname{arc\,tg} \frac{1}{\sqrt{(m+n)-1}} \right]. \tag{21}$$

For a given value of m we can, by a graphic solution of equation (21), find the value of the coefficient n for which A = 1 (Fig. 6).

Thus, for m = 1, A = 1 for n ≈ 5, for m = 1.1, the condition A = 1 holds for n ≈ 13.2, etc.

In the second case, i.e., for determining the boundary bc (Fig. 5):

$$A = \frac{b^2 + \omega_1^2}{b^2} \frac{q_2^{\frac{2q_2}{q_1 - q_2}}}{q_1^{\frac{2q_1}{q_1 - q_2}}} \exp \frac{2b}{\omega_1} \left[-\frac{\pi}{2} - \operatorname{arc\,tg} \frac{b}{\omega_1} \right]. \tag{22}$$

where

$$\omega_1 = \sqrt{(\omega_0 + \beta) - b^2}. \quad q_1 = -b + \sqrt{b^2 - \omega_0}. \quad q_2 = -b - \sqrt{b^2 - \omega_0}.$$

By replacing the values of ω_0 and β by mb^2 and nb^2, we get

$$A = (m+n) \frac{(-1 - \sqrt{1-m})^{\frac{-1-\sqrt{1-m}}{\sqrt{1-m}}}}{(-1 - \sqrt{1-m})^{\frac{-1+\sqrt{1-m}}{\sqrt{1-m}}}} \times$$

$$\times \exp \frac{2}{\sqrt{(m+n)-1}} \left[-\frac{\pi}{2} - \operatorname{arc\,tg} \frac{1}{\sqrt{(m+n)-1}} \right]. \tag{23}$$

By determining, in an analogous way, the values of the coefficients m and n for the case A = 1, we find that n = 5 for m = 1, n = 15 for m = 0. By placing these points of the boundary of stability on the ω_0, β plane, we determine the region of stability for the case $J_j = 0$. The stability region is broader than in the case when there was no self-smoothing.

The stable portion of the plane of the parameters ω_0 and β falls into a number of regions which are distinguished by their different indicators of system quality.

Region 1 (Fig. 5) is the region of oscillatory processes, since the motion of the representative point on sheet I, for $\omega_0 - \beta > b^2$, is defined by the equation

$$(\dot{\varphi} + b\varphi)^2 + \omega_1'^2 \varphi^2 = C \exp \frac{2b}{\omega_1'} \operatorname{arctg} \frac{\dot{\varphi} + b\varphi}{\omega_1' \varphi}. \tag{24}$$

where $\omega_1' = \sqrt{(\omega_0 + \beta) - b^2}$.

On sheet II, the motion of the representative point is defined either by equation (24) or by the equation

$$(\dot{\varphi} + q_1\varphi)^{q_1} = c(\dot{\varphi} + q_2\varphi)^{q_2}. \tag{25}$$

where $q_1 = -b + \sqrt{b^2 - \omega_0}$ and $q_2 = -b - \sqrt{b^2 - \omega_0}$.

Thus, the phase plane in this region is filled with trajectories consisting of spirals with different values of the coefficient ω_1^2, or of combinations of trajectories of the parabolic type with spirals.

Region 2 is included between the lines $\omega_0 - \beta = b^2$ and $\omega_0 = \beta$, and is the region of processes without overshoot. The trajectories filling the phase plane for $\omega_0 < b^2$ consist of curves of the parabolic type, defined by equation (25), where the values of the roots q_1 and q_2, for the plane of sheet I, equal

$$q_1 = -b + \sqrt{b^2 - (\omega_0 \pm \beta)}, \quad q_2 = -b - \sqrt{b^2 - (\omega_0 \pm \beta)},$$

and, for the plane of sheet II, equal

$$q_1 = -b + \sqrt{b^2 - \omega_0}, \quad q_2 = -b - \sqrt{b^2 - \omega_0}.$$

For $\omega_0 > b^2$, the phase plane will be filled with trajectories consisting of segments of spirals and curves of the parabolic type.

Region 3 will correspond to a "sticking" of the representative point on the φ axis. Here, as for the case when self-smoothing was lacking, the equation $\varphi_{st} = f(\varphi_{II}\dot{\varphi}_I)$ is valid.

We note that, in choosing the working point in region 2 for $J_j = 0$, it is impossible to obtain any serious gain in the duration of the transient response in comparison with the case of system stabilization by the introduction of ordinary derivatives. It is therefore advantageous to consider the deformation of the region of stability when the value of the coefficient J_j is varied. This is necessary in estimating the roughness of the system and in choosing the working point on the ω_0, β plane with the best transient response quality.

As the coefficient J_j varies from 0 to ∞, one of the stability boundaries, namely, $\omega_0 = 0$, will undergo no change. The other boundary, abc (Fig. 5), will be translated toward the region of more negative values of β and, for $J_j = \infty$, will vanish at infinity. (Fig. 7). Analysis of the phase plane shows that the highest indicators of quality correspond to regions 2 and 3 of the ω_0, β plane with $\omega_0 > b^2$, for the case $\infty > J_j > 0$.

The Cross Section for $\infty > J_j > 0$ ($k_j \neq 0$, $T_j \neq 0$)

As in the case of no self-smoothing, the motion of the representative point will proceed on a two-sheeted phase plane.

The boundaries of the sheets are defined by the equations $\varphi = 0$, $\dot{\varphi} = -J_j\varphi$, i.e., as in the previous case, the presence of the coefficient J_j increases the area of sheet II and decreases the area of sheet I.

In this case, region 2 falls into two regions, 2a and 2b (Fig. 8).

Region 2a ($-b- \sqrt{b^2 - (\omega_0 - \beta)} < -J_j$) is characterized by the fact that the asymptote, q_2, lies in the plane of sheet II and, thanks to this, a damped oscillatory process occurs. If the asymptote q_2 lies in the plane of sheet I or on the boundary between the sheets (region 2b, $-b- \sqrt{b^2 - (\omega_0 - \beta)} \geq -J_j$), the transient response will occur without overshoot where, if the asymptote q_2 coincides with the boundary between the sheets, $\dot{\varphi} = -J_j\varphi$, the relay element will transfer only once, and the representative point will approach the origin of coordinates without sliding.

It should be mentioned that, in this case, the motion of the representative point will proceed on the two-sheeted phase plane, filled by trajectories consisting of segments of spirals (sheet II) and curves of a parabolic type (sheet I). In this case, at the end of the transient response the system, by its structure, will be stable and, therefore, there will be no autooscillations about the equilibrium position engendered by dead zones and the hysteresis loop of the relay element.

Region 3 is characterized by a change of the equation of motion of the representative point on sheet I. The plane of the sheet will be filled with families of shifted hyperbolae with asymptotes

$$q_1 = -b + \sqrt{b^2 + (\beta - \omega_0)} \ ; \ q_2 = -b - \sqrt{b^2 + (\beta - \omega_0)}.$$

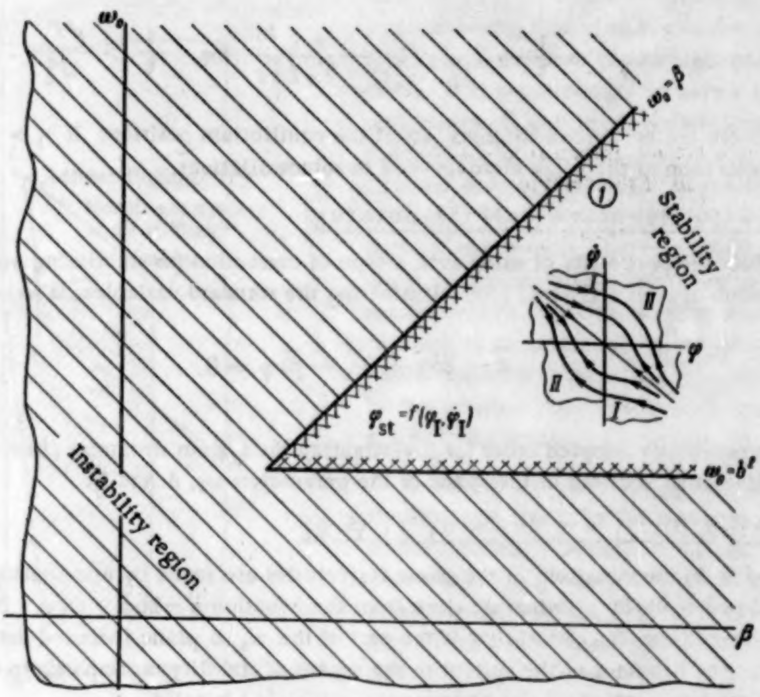


Fig. 9. Plane of the parameters ω_0 and β for $J_j = 0$.

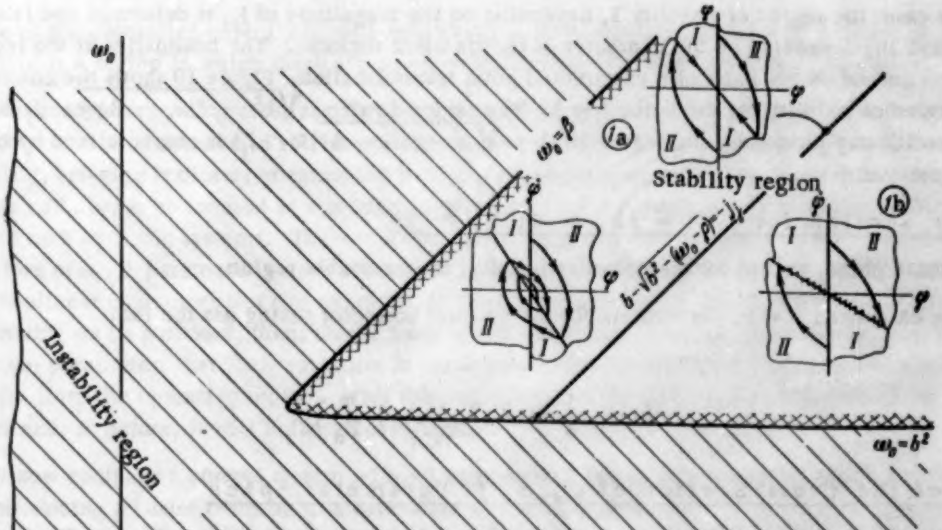


Fig. 10. Plane of the parameters ω_0 and β for $\infty > J_j > 0$ ($J_j = 1$).

This region is also divided into two regions: region 3a ($-b - \sqrt{b^2 + (\beta - \omega_0)} < -J_j$) where oscillatory processes occur and region 3b ($-b - \sqrt{b^2 + (\beta - \omega_0)} > -J_j$) in which the transient response will not have overshoot. Just as for the previous case, when the asymptote q_2 coincides with the boundary between the sheets, $\dot{\varphi} = -J_j\varphi$, the representative point approaches the origin of coordinates without sliding.

On the basis of the analysis given, one may make recommendations as to the tuning of the correcting device for the case when the object has positive self-smoothing. The formula for optimal tuning has the form

$$k_i = \frac{1}{\delta} + T_a T_s \left(\frac{k_j}{T_j} - \frac{p}{T_a}\right)\frac{k_j}{T_j} \quad \text{for} \quad \frac{1}{\delta} > \frac{p^2 T_s}{4T_a^2}. \tag{26}$$

If $k_i < 1/\delta$, there are no autooscillations about the equilibrium position. If $k_i > 1/\delta$, then the dead zone and the hysteresis loop of the relay element lead to autooscillations.

3. An Object with Negative Self-Smoothing

We now consider the possibility of using such a type of correction for stabilizing an unstable object ($\rho < 0$). The equation of motion of such a system, after eliminating the standard variables, is written in the form

$$\ddot{\varphi} - 2b\dot{\varphi} + (\omega_0 - \bar{\beta})\varphi = 0. \tag{27}$$

Following the previously adopted order for investigating the system dynamics (free oscillations), we consider the corresponding cross sections of the space of the parameters ω_0, β and J_j.

The Cross Section for $J_j = 0$ ($k_j = 0$, $T_j \neq 0$)

On sheet I ($\varphi\dot{\varphi} < 0$) the equations of the phase trajectories are found by eliminating time from equation (27) and, on sheet II ($\varphi\dot{\varphi} > 0$), by eliminating time from the equation $\ddot{\varphi} - 2b\dot{\varphi} + \omega_0\varphi = 0$. Investigation of the phase plane showed that the region of stability is the part of the ω_0, β plane included between the lines $\omega_0 = \beta$ and $\omega_0 = b^2$ (Fig. 9). The behavior of the system in the region of stability corresponds to the case of the "sticking" representative point on the φ axis, i.e., the equation $\varphi_{st} = f(\varphi_I, \dot{\varphi}_I)$ is valid.

The Cross Section for $\infty > J_j > 0$ ($k_j \neq 0$, $T_j \neq 0$)

In this case, the region of stability 1, depending on the magnitude of J_j, is deformed and falls into two regions, 1a and 1b, depending on the character of the transient response. The boundaries of the region of stability are determined on the basis of a constructed point transformation. Figure 10 shows the constructed boundary of the region of stability for the value $J_j = 1$. The region 1a ($b - \sqrt{b^2 - (\omega_0 - \beta)} < -J_j$ is characterized by damped oscillatory processes, and region 1b ($b - \sqrt{b^2 - (\omega_0 - \beta)} \geq -J_j$) is characterized by transient responses without overshoot.

The Cross Section for $J_j = \infty$ ($k_j = 0$, $T_j = 0$)

The whole plane, except for the points at infinity, is an unstable region.

For the case when $\rho < 0$, the formula for the optimal corrector tuning has the form

$$k_i = \frac{1}{\delta} + T_a T_s \left(\frac{k_j}{T_j} + \frac{p}{T_a}\right)\frac{k_j}{T_j}. \tag{28}$$

Analysis of the Cross Section of ω_0, β, J_j Parameter Space

In analyzing the corresponding cross sections of ω_0, β, J_j parameter space, one may draw certain conclusions with respect to stability, transient response quality and also the efficacy of employing such a type of corrector, as compared to system stabilization by means of introducing a linear response to derivatives in the control law. It is also possible to answer the questions of roughness of such dynamic systems, of the effect of the nonlinear characteristics of the object on the quality of the transient response and also to determine the moments of time at which it is necessary the vary the system's structure in order to obtain high-quality transient responses.

The formulae for optimal tuning of the correcting device bespeak the possibility of obtaining high-quality systems with the presence of one Ψ-cell, the tuning of which is done as a function of the object and controller parameters in accordance with the formulae given above.

It is clear, from an analysis of these formulae and the phase planes corresponding to them, that it is sufficient, to decrease the duration of the transient response, to increase the gain with respect to the controlled

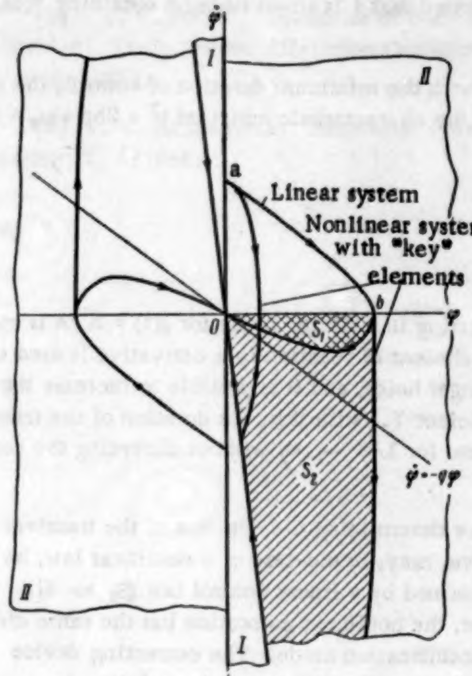


Fig. 11

coordinate, $1/\delta$, with corresponding corrections in the optimal tuning formulae due to the coefficients k_j and J_j. Consequently, to obtain high-quality control processes, it is not necessary to have a large value for the coefficient of reaction to a derivative T_j, since it is required to decrease the coefficient T_j ($J_j = k_j/T_j$) to increase the coefficient J_j. In practice, the magnitude of T_j will depend on the insensitivity of the relay element of the Ψ-cell. Analysis of the corresponding ω_0, β parameter planes shows that a change of system structure from stable to unstable is possible at any moment of time after the maximum value of the controlled quantity is attained; thus the control time will be the smaller, the more remote from this point is the act of transferring.

At the point $\varphi = \varphi_{max}$ it is impossible to vary the structure since, in this case, the steady-state value of the controlled coordinate will depend on the initial conditions. At this point, to improve the quality of the control process, it is possible only to decrease the velocity of the controlling organ. Since, at the end of the transient response, the system will be unstable, about the equilibrium position, due to the dead zone of the relay element, there will arise autooscillations, the amplitudes and frequencies of which depend on the system parameters.

The case of the controlled object with self-smoothing is an exception. In this case, on the ω_0, β parameter plane there is a region in which there are no auto-oscillations about the equilibrium position.

Analysis of the corresponding cross-sections of the ω_0, β, J_j parameter spaces shows that the structure of the object – the presence or absence of self-smoothing (positive or negative) – deforms the boundary of the region of stability, bringing it closer, or removing it from, the working (operating) region. However, the chosen operating points can always be located at a significant distance from the boundary of stability, which speaks for the roughness of such dynamic systems. The conditions for the existence of processes without overshoot show that entire regions of ω_0, β parameter planes correspond to such processes. Consequently, if the controlled object possesses nonlinear characteristics (for example, if there is a variation in the effectiveness of the controlling organ depending on its position), then, on the basis of the theoretical analysis given above, one may so choose the system parameters that their variation in accordance with the nonlinear characteristics will not remove the system from the operating region. With this, no change in the qualitative character of the transient response occurs and, as before, it will not have overshoot.

We shall now compare a control system with a linear control law with a system in which stabilization is obtained by introducing a linear reaction to a derivative in the control law.

For simplicity, we choose a controlled object without self-smoothing. After elimation of variables, the equation of the control system is written in the form

$$\ddot{\varphi} + 2b\dot{\varphi} + \omega_0\varphi = 0, \tag{29}$$

where

$$2b = \frac{T}{T_a T_s}, \quad \omega_0 = \frac{1}{T_a T_s \delta},$$

and T is the coefficient of the reaction to the derivative.

Let T equal some sufficiently small constant, $\underline{i}$. It is assumed that $\underline{i}$ is small because obtaining "clean" (undistorted) derivatives with large gains is quite difficult.

Then, to obtain a transient response without overshoot and with the minimum duration of control, the size of the gain, $1/\delta$, is determined by the condition that the roots of the characteristic equation $p^2 + 2bp + \omega_0 = 0$ be equal ($q_1 = q_2 = q$).

$$\frac{1}{\delta} = \frac{i^2}{4T_a T_s}. \tag{30}$$

In this case, the minimum control time for the system operating in a sliding mode for $g(t) = A$ (A is some constant) is determined by the area S_1 (Fig. 11). If this same coefficient of reaction to a derivative is used to form a nonlinear law, $\Psi_{ij}(\varphi, \dot{\varphi})\,\varphi(t)$, the function in (30) no longer holds, and it is possible to increase the gain without limit, without a corresponding increase in the coefficient T. With this, the duration of the transient response will decrease to a quite small value (theoretically, to zero for $1/\delta \to \infty$) without disrupting the condition for a transient response without overshoot.

It is clear from Fig. 11 that the area of the subintegral curve determining the duration of the transient response, for one and the same coefficient of response to a derivative, may, in the case of a nonlinear law, be made significantly larger than the area of the subintegral curve defined by a linear control law [$S_2 \gg S_1$, $t(S_2) \ll t(S_1)$]. One may easily convince oneself that, in this case, the nonlinear correction has the same effectiveness whether the system operates in a sliding mode or in the stabilization mode. The correcting device does not require readjustment to carry out high-quality control processing for different points of disturbance application.

Theoretically, for the gain $1/\delta \to \infty$ ($J_j \to \infty$), when the system operates in the sliding mode, the dynamic error tends to zero, $S_2 \to \infty$ (Fig. 11), i.e., the system reproduces quite accurately any disturbing stimulus g (t) down to skips. When the system operates in the stabilization mode (for $1/\delta \to \infty$), the duration of the transient response tends to zero, i.e., the system reacts weakly to any disturbance.

It should be mentioned that similar results can be obtained in control systems with linear reactions to derivatives if the gain $1/\delta$ is increased to ∞^2 and the coefficient of reaction to the derivative, T, is simultaneously increased to ∞.

Consequently, if a nonlinear correcting device is employed, the number of high-gain amplifiers can be halved, since the absolute value of the gain, $1/\delta$, for obtaining high-quality control processes must be less than in a linear system. Thus, to obtain a regulation time $t \to 0$, in a nonlinear system $1/\delta \to \infty^2$.

The author wishes to express his appreciation to B. N. Petrov under whose direction the work described here was carried out.

SUMMARY

The paper deals with the problem of stabilizing and improving the quality of the second-order automatic control systems by means of nonlinear compensation "key"-type devices.

LITERATURE CITED

[1] S. V. Emel'yanov, "A method of obtaining complex control laws using only an error signal or the coordinate to be controlled and its first derivative," [in Russian]. Automation and Remote Control (USSR) 18, 10 (1957).

[2] A. A. Andronov and S. E. Khaikin, Theory of Oscillations. [in Russian]. Gostekhizdat (1937).

[3] V. V. Petrov, "On the autooscillations of two-stage servomechanisms with relay controls," [in Russian]. Automation and Remote Control (USSR) 12, 1 (1951).

[4] V. V. Petrov and G. M. Ulanov, "The theory of two simplest relay systems for autocontrol," [in Russian]. Automation and Remote Control (USSR) 11, 5 (1950).

[5] V. V. Petrov, "Dynamics of one- and two-stage servomechanisms with nonlinear characteristics," [in Russian]. Trudy Second All-Union Conference on Automatic Control Theory, Vol. [in Russian]. Izdatelstvo AN SSSR (1955).

[6] V. V. Kazakevich, "Sequential systems and the simplest dynamic models of time," [in Russian]. Dokl. AN SSSR 74, 4 (1950).

Received October 27, 1958

OPTIMAL LAWS FOR ELECTRIC-DRIVE CONTROL

Yu. P. Petrov

(Leningrad)

The paper considers optimal laws for electric-drive control which provide maximum productivity for given limitations on heating. Optimal control laws are derived for various forms of functional dependences of the motor's magnetic flux on the armature current and of the impedance moment on the rotational speed.

In many electric drives, particularly those working in motor-generator systems, there exists the capability of continuously varying the current flowing in the armature of the executive electric drive. Therefore, the problem of finding an optimal law for varying this current is of interest. This law, as will be shown below, depends primarily on the requirements placed on the electric drive.

There are two basic types of requirements. The requirements of the first type amount to the demand that the greatest speed variation occur in a given time for a given heat loss in the armature, corresponding to the total heat used by the motor. The second type of requirement consists of the demand that a given armature heat loss provide the greatest displacement of the executive mechanism in a given time.

The requirements of the first type apply to the electric drives of mechanisms whose operating processes require completely definite rotational speeds. Reaching of the "operating" speed must be accomplished in the least possible time. Examples of such electric drives are the starting motors of synchronous compensators, centrifuges, propeller motors, etc.

The requirements of the second type are applied to the auxiliary mechanisms of rolling mills, to traction motors, to the electric drives of hoists, to artillery stands, etc.

We introduce the following notation: $i = I/I_N$ is the relative magnitude of the armature current, where I_N is the nominal current; $\mu = M_I/M_N$ is the relative magnitude of theimpedance moment, where M_N is the nominal moment; $\nu = n/n_N$ is the relative magnitude of the rotation speed, where n_N is the nominal speed; $\Phi = \Phi_{mo}/\Phi_N$ is the relative magnitude of the useful resulting magnetic flux of the motor, where Φ_N is the nominal flux; $\tau = t/T_M$ is the relative time, where $T_M = CD^2 n_N / 375 M_N$ is the electromechanical time constant.

For any dc electric drive we may write

$$i\Phi = \frac{d\nu}{d\tau} + \mu, \tag{1}$$

where, generally speaking, the magnetic flux is a function of the current. The quantity of heat, ΔQ, given off by the armature during the time $\Delta\tau = \tau_2 - \tau_1$ is expressed by the integral

$$\Delta Q = A\int_{\tau_1}^{\tau_2} i^2 d\tau, \tag{2}$$

where A is a constant.

The increment of speed during time $\Delta\tau$ is expressed by the integral

$$\Delta v = \int_{v_1}^{v_2} dv, \tag{3}$$

and the increment of the angle rotated through is given by the integral

$$\Delta\alpha = \int_{\tau_1}^{\tau_2} v d\tau. \tag{4}$$

We now seek the optimal law for controlling the current through the armature for an electric drive subject to a requirement of the first type. The magnetic flux is a function of current, while we assume the impedance moment to be constant for the time being.

On the basis of equation (1) we may transform the integral in (3) to the form

$$\Delta v = \int_{\tau_1}^{\tau_2} (i\Phi - \mu)\, d\tau. \tag{5}$$

The problem of optimal control is formulated mathematically as follows: it is required to find such a function, $i = i(\tau)$, which, for a given magnitude of the integral in (2), would make the integral in (5) a maximum. The solution of this isoperimetric problem from the calulus of variations we find from Euler's equation

$$\frac{\partial F}{\partial i} - \frac{d}{d\tau}\frac{\partial F}{\partial i'} = 0$$

for $F = i^2 - \lambda(i\Phi - \mu)$, where λ is a constant.

By differentiating, we get

$$2i - \lambda\left(\Phi + \frac{d\Phi}{di} i\right) = 0,$$

from which it follows that the optimal control law is to maintain a constancy of the current: $i_{opt} = c$, where $\underline{c}$ is the root of equation (6). An exception is the case when the magnetic flux is proportional to the current: $\Phi = ki$. In this case, $d\nu/d\tau = ki^2 - \mu$ and, consequently, $\Delta\nu = k\Delta Q - \mu\Delta\tau$, i.e., the change of speed is completely determined by giving $\Delta\tau$ and ΔQ, and does not depend on the law of current variation. For a series unsaturated motor, where $\Phi = ki$, any control law is as optimal as any other.

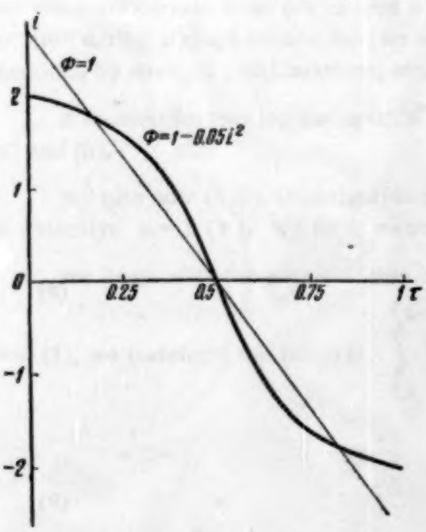


Fig. 1

We turn now to electric drives to which requirements of the second type are applied.

The optimal control is formulated mathematically as follows: it is required to find $i = i(\tau)$ and $\nu = \nu(\tau)$ for which the integral in (4) attains a maximum for a given magnitude of the integral in (2) and a given relationship equation (1). We find the solution of this general problem from Euler's equation

$$\frac{\partial F}{\partial i} - \frac{d}{d\tau}\frac{\partial F}{\partial i'} = 0, \quad \frac{\partial F}{\partial v} - \frac{d}{d\tau}\frac{\partial F}{\partial v'} = 0 \tag{7}$$

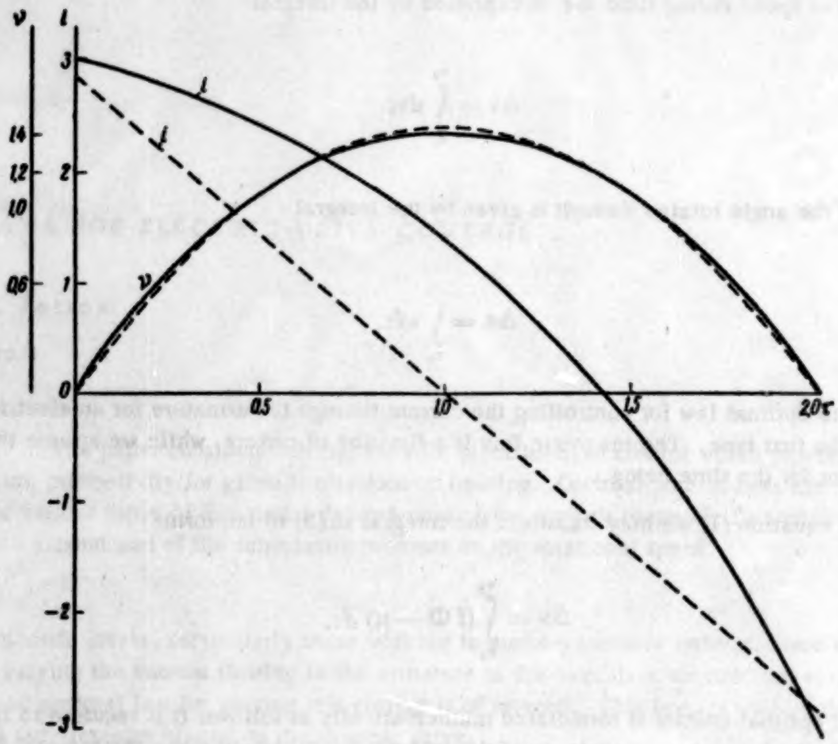


Fig. 2

for

$$F = i^2 + 2\lambda\nu + 2\lambda_1(\nu' - i\Phi + \mu).$$

Here λ is a constant and λ_1 is a function of [1].

By differentiating, we get

$$2i - 2\lambda_1\left(\Phi + i\frac{d\Phi}{di}\right) = 0. \quad \frac{d\lambda_1}{d\tau} = \lambda.$$

from which

$$i = (\lambda\tau + C)\left(\Phi + i\frac{d\Phi}{di}\right). \tag{8}$$

Equation (8) defines the optimal current curve.

We now consider the particular case

$$\Phi = 1 - ai^2. \tag{9}$$

corresponding to an independently excited motor, in which the demagnetizing action of the armature reaction may be considered to follow equation (9) approximately. The magnitude of the coefficient a, for noncompensated motors, is generally found between 0.05 and 0.1. In the given case, we write equation (8) in the form

$$i = (\lambda\tau + C)(1 - 3ai^2).$$

which gives

$$i = \frac{-1 \pm \sqrt{1 + 12a(\lambda\tau + C)^2}}{6a(\lambda\tau + C)}. \tag{10}$$

The constants λ and C are determined from the condition $\int_{\tau_1}^{\tau_2} i^2 d\tau = Q$ and from the boundary conditions: $\nu = \nu_1$ for $\tau = \tau_1$ and $\nu = \nu_2$ $\tau = \tau_2$. By setting a = 0, i.e., by ignoring the demagnetizing action of the armature reaction, we obtain the particular case of a constant magnetic flux. The optimal control law for this case may be obtained from (10) by going to the limit as $a \to 0$:

$$i = \lambda\tau + C \tag{11}$$

This particular case was investigated in detail by K. I. Kozhevnikov [2] and E. A. Rozenman [3].

In order to show graphically to what degree the armature reaction influences the deviation of the optimal current curve from the linear one, defined by equation (11). Fig. 1 shows the optimal current curve for $\Phi = 1 - 0.05\, i^2$, $\nu_1 = \nu_2 = 0$, $\mu = 0$, and the optimal curve for $\Phi = 1$, equivalent to the first in terms of heating.

According to the duality principle [1], the optimal control guarantees, not only the maximum translation or speed variation for a given heat loss, but also the least heat loss for a given speed change or change in angle rotated.

Consequently, the requirements placed on electric drive may be formulated in the following manner. A requirement of the first type: for a given change of speed in a given time it is necessary to provide the minimum heat loss in the armature. A requirement of the second type: for a given translation of the executive mechanism in a given time, it is necessary to provide the minimum heat loss in the armature. In these cases, the optimal laws of current variation will be expressed, respectively, by equations (6) and (10).

Of course, in addition to the limitations on heat mode, other limitations may be placed on a motor. Thus, the armature current must not exceed a limiting value permitted by the commutation conditions; the speed of rotation during a given translation (for requirements of the second type) must not exceed a limiting value determined by strength considerations, etc.

A method for seeking the optimal current when these additional factors are taken into account is given in [2] and [5].

We turn now to the investigation of optimal control laws for the drive's impedance moment as a function of velocity: $\mu = \mu(\nu)$. We limit ourselves to the case of constant magnetic flux, $\Phi = 1$.

We begin with the consideration of laws which satisfy requirements of the first type. On the basis of equation (1), we transform the integral $\int_{\tau_1}^{\tau_2} i^2 d\tau$ to the form

$$\int_{\tau_1}^{\tau_2} (\nu' + \mu)^2 \, d\tau. \tag{12}$$

and the integral $\int_{\nu_1}^{\nu_2} d\nu$ to the form

$$\int_{\tau_1}^{\tau_2} \nu' d\tau. \tag{13}$$

After this, the problem of optimal control reduces to the isoperimetric problem of the calculus of variations: it is required to find such a function, $\nu = \nu\ (\tau)$, for which, with a given value of the integral in (12), the integral in (13) assumes a maximum value. We find the solution of this problem from Euler's equation

$$\frac{\partial F}{\partial \nu} - \frac{d}{d\tau}\frac{\partial F}{\partial \nu'} = 0.$$

for

$$F = \lambda \nu' + (\nu' + \mu)^2.$$

By differentiating we find that

$$\frac{\partial F}{\partial \nu} = 2\nu' \frac{d\mu}{d\nu} + 2\mu \frac{d\mu}{d\nu}.$$

$$\frac{\partial F}{\partial \nu'} = \lambda + 2\nu' + 2\mu.$$

$$\frac{d}{d\tau}\frac{\partial F}{\partial \nu'} = 2\nu'' + 2\frac{d\mu}{d\tau}.$$

By taking into account the fact that $d\mu/d\tau = (d\mu/d\nu)(d\nu/d\tau)$, we put the Euler equation into the form

$$\frac{d^2\nu}{d\tau^2} - \mu \frac{d\mu}{d\nu} = 0. \tag{14}$$

If we multiply all the terms of (14) by $2\nu' d\tau = 2d\nu$ and integrate, we get

$$\left(\frac{d\nu}{d\tau}\right)^2 = \mu^2 + C. \tag{15}$$

By taking account of (1), we can find from equation (15) the relationship between current and velocity: $\nu' = i - \mu$.

Whence, $(1 - \mu)^2 = \mu^2 + C$ and, consequently,

$$i = \mu \pm \sqrt{\mu^2 + C}. \tag{16}$$

The plus sign refers to the case of motor starting, and the minus sign to the case of braking.

We now consider in more detail the special case when $\mu = k\nu$. In this case, the Euler equation (14) takes the form: $\nu'' - k^2\nu = 0$, whence

$$\nu = C_1 e^{k\tau} + C_2 e^{-k\tau}. \tag{17}$$

$$i = 2kC_1 e^{k\tau}. \tag{18}$$

Consequently,

$$Q = \int_0^{\tau_0} i^2 d\tau = 2kC_1^2 (e^{2k\tau_0} - 1). \tag{19}$$

For a given Q, we find the magnitude of C_1 from equation (19) and then, knowing the magnitude of ν_0, from equation (17) we find C_2.

We consider an example. Let $k = 1, \nu = 0, \tau = 1$ and $\int_0^1 i^2 d\tau = 3.19$. Then, $i = e^{\tau}, \nu = (e^{\tau} - e^{-\tau})/2$ and for $\tau = 1$, we have that $\nu = 1.17$.

Let us now assume that the magnitude of the current is kept constant. In this case, with the same value for the integral $\int_0^1 i^2 d\tau = 3.19$, i.e., with the same degree of motor heating, $i - 1.79, \nu = 1.79(1 - e^{-\tau})$ and, for $\tau = 1, \nu = 1.13$. In other words, the change of speed will be approximately 4% less than with the optimal control law. Thus, in contradistinction to a constant impedance moment, an impedance moment which depends on the velocity significantly affects the form of the optimal current curve.

The magnitude of the constant C in formula (16) is determined by giving the interval of time during which starting of the motor must be completed. If the starting time is not given, but it is required to provide the minimum possible heat loss, then in this case $C = 0$. Indeed, if we replace the variables $\underline{i}$ and τ by μ and ν in the integral in (2), we then obtain, in accordance with equations (15) and (16),

$$Q = \int_{\nu_1}^{\nu_2} \frac{(\mu \pm \sqrt{\mu^2 + C})^2}{\sqrt{\mu^2 + C}} d\nu.$$

Whence, differentiating with respect to C and equating the derivative to zero, we get $Q = Q_{min}$ for $C = 0$, i.e., when $i = \mu \pm \mu$. For the case of starting we have that $i = 2\mu$ and for the case of braking, the trivial solution, $i = 0$. This last equality means that the minimum heat loss, namely a zero loss, occurs in the case when the drive is braked solely due to its inherent impedance moment.

We now turn to the consideration of electric drives subject to requirements of the second type. In this case, the problem of finding the optimal control law is formulated as follows: it is required to find such a function, $\nu = \nu(\tau)$, that, for a given value of the integral in (12), the integral in (3) will be maximized. Thus, we are again led to the isoperimetric problem of the variational calculus, the solution of which we find from Euler's equation

$$\frac{\partial F}{\partial \nu} - \frac{d}{d\tau}\frac{\partial F}{\partial \nu'} = 0.$$

for $F = (\nu' + \mu)^2 - 2\lambda\nu$.

By differentiating, we obtain Euler's equation in the form

$$\frac{d^2\nu}{d\tau^2} - \mu\frac{d\mu}{d\nu} + \lambda = 0. \tag{20}$$

By multiplying all the terms of (20) by $2\nu' d\tau = 2d\nu$ and integrating, we get

$$\left(\frac{d\nu}{d\tau}\right)^2 = \mu^2 - 2\lambda\nu + C. \tag{21}$$

From this, by taking (1) into account, we obtain the relationship between current and speed:

$$i = \mu \pm \sqrt{\mu^2 + C - 2\lambda\nu}. \tag{22}$$

The plus sign is taken for that portion of the complete translation cycle when starting of the drive occurs, and the minus sign is taken for the braking portion.

We now consider in detail the case when $\mu = a + k\nu$. For the given case, the Euler equation (20) takes the form:

$$\nu'' - k^2\nu + \lambda - ka = 0. \tag{23}$$

From this, by letting $\lambda - ka = m$, we get

$$\nu = C_1 e^{k\tau} + C_2 e^{-k\tau} + \frac{m}{k^2}, \quad i = 2kC_1 e^{k\tau} + \frac{m}{k} + a.$$

From the boundary conditions, $\nu = 0$ for $\tau = 0$ and $\nu = 0$ for $\tau = \tau_0$, we determine the constants C_1 and C_2:

$$C_1 = -\frac{m}{k^2}\left(\frac{1 - e^{-k\tau_0}}{e^{k\tau_0} - e^{-k\tau_0}}\right), \quad C_2 = -\frac{m}{k^2}\left(\frac{e^{k\tau_0} - 1}{e^{k\tau_0} - e^{-k\tau_0}}\right).$$

The constant $\underline{m}$ is determined by the condition that $\int_0^{\tau_0} i^2 d\tau = Q$.

As an example, Fig. 2 shows the curves for current and speed for the case $a = 0$, $k = 1$, $\lambda = 4$ and $\tau_0 = 2$.

These curves are described by the equations

$$\nu = 4 - 0.476\, e^{\tau} - 3.524\, e^{-\tau}. \quad i = 4 - 0{,}952\, e^{\tau}.$$

For purposes of comparison, the dotted lines on Fig. 2 show the optimal curves for zero impedance moment and for the same value of angular rotation during time τ_0. The dotted curves are described by the equations

$$\nu = 2.88\tau - 1.44\tau^2. \quad i = 2{,}88 - 2.88\tau.$$

For $\mu = \text{const}$, we obtain from (20) and (22) the well-known [2, 3] special case of the optimal control law by which the armature current is a linear function of ti e and the speed is a parabolic function of time. The example given shows graphically that, for $\mu \neq$ const, the current law differs significantly from the linear, while the velocity law differs only insignificantly from a parabolic one.

The problem of the optimal control law when voltage is controlled, in its usual setting, when the motor's magentic flux depends on current and the impedance moment depends on speed, also reduces to the general Lagrange problem in the calculus of variations.

Thus, for example, if it is necessary to meet a requirement of the first type, then the finding of the optimal law of current regulation reduces to integrating the Euler equations

$$\frac{\partial H}{\partial i} - \frac{d}{d\tau}\frac{\partial H}{\partial i'} = 0. \quad \frac{\partial H}{\partial \nu} - \frac{d}{d\tau}\frac{\partial H}{\partial \nu'} = 0.$$

where $H = i^2 + \lambda\varphi$, with λ a function of τ, and

$$\varphi = i\Phi - \frac{d\nu}{d\tau} - \mu.$$

By differentiating we obtain

$$2i + \lambda\left(\Phi + \frac{d\Phi}{di}i\right) = 0. \quad \frac{d\lambda}{d\tau} = \lambda\frac{d\mu}{d\nu}. \tag{24}$$

Equations (24), together with the condition that $\varphi = 0$, suffice to define the three unknowns, i, ν, λ.

Let us integrate (24) for the case when $\mu = m + k\nu$, $\Phi = 1 - ai^2$. In this case, $d\mu/d\nu = k$, $\Phi + i d\Phi/di = 1 - 3ai^2$. We find, from the second of equations (24) that $\lambda = Ce^{k\tau}$. By substituting for λ in the first of equations (24), we get

$$i = \frac{1 \pm \sqrt{1 + 3a\, C^2 e^{2k\tau}}}{3a\, Ce^{k\tau}}.$$

By letting $k = 0$ or $a = 0$, we obtain, respectively, $i = \text{const}$ [Cf. (6)] or $i = -Ce^{k\tau}/2$ [Cf. (18)].

It is possible to seek an optimal control law while taking account of the heat transfer from the motor to the ambient space. In this case, the limiting conditions will not depend on the amount of heat evolved in the winding, but will depend on the maximum overheating during the operating cycle [3, 4]. However, in the majority of cases, the motor's heating constant is many times larger that the time of the operating cycle, so that taking the heat transfer into account affects the control law only insignificantly [4].

Of course, implementing the optimal control laws in practice is not always possible. The requirement of providing maximum productivity of the electric drive may run counter to the requirements of simplicity and reliability of the control devices. In any case, knowing the optimal control laws allows one to find the maximum possible productivity of an electric drive and to determine to what degree some actual control system approximates to the optimal.

SUMMARY

1. The optimal control law depends on the type of requirement placed on the electric drive. In the majority of engineering problems, these requirements reduce either to a demand for the maximum variation in speed or to the demand for the maximum translation of the mechanism in a given time.

2. The choice of the optimal control law is determined by the forms of the relationship of the magnetic flux to the current, and the impedance moment to the speed, and by the type of requirement placed on the electric drive. Each case has its own optimal control law.

The "right-angled current curve" with charging coefficient equal to unity, and the linear current curve [2, 3] are special cases of these optimal laws.

LITERATURE CITED

[1] M. A. Lavrent'ev and L. A. Lyusternik. Course in Variational Calculus. [in Russian]. Gostekhizdat (1950).

[2] K. I. Kozhevnikov, "Motor current curve for auxiliary mechanisms of rolling mills," [in Russian]. Élektrichestvo, 6 (1956).

[3] E. A. Rozenman, "On the optimal transient responses in systems with power limitations," [in Russian]. Automation and Remote Control (USSR) 18, 6 (1957).

[4] Sh. Sh. Khamitov, "An investigation of dc motors as objects of optimal control systems," [in Russian]. Élektrichestvo 5 (1958).

[5] E. A. Rozenman, "On the limiting speeds of action of servo systems limited in power, speed and torque of the executive mechanism," [in Russian]. Automation and Remote Control (USSR) 19, 7 (1958).

Received August 7, 1958

THE REPRODUCTION OF SINUSOIDAL SIGNALS BY SYSTEMS WITH LIMITATIONS

G. A. Nadzhafova
(Moscow)

Considered in this paper are the conditions for partial, and complete, irreproducibility of sinusoidal signals, i.e., the limiting frequency characteristics are determined for standard second-order systems consisting of linear astatic, or static, objects controlled by an astatic servomotor which is limited as to the speed of action and the course of the controlling organ. The amplitude errors of reproduction are determined.

INTRODUCTION

Many objects in automated productive processes may be approximately characterized by linear first-order differential equations. To control the supply of energy or material to these objects one uses, as executive mechanisms, power drives which, by their properties, are close to linear integrating servomotors. For such systems there are always limitations on the speed of operation of the servomotor and on the movement of the controlling organ. Due to the presence of these limitations, it is impossible to improve indefinitely the system's speed of response, or other dynamic indicators of the transient response or steady-state processes.

In [2] there were investigated the questions related to the determination of the limiting dynamic indicators of the transient responses occurring in such automatic control systems.

However, in many practical cases of automated productive processes as, for example, in programmed-control and servo systems, there are important questions connected with the passage of sinusoidal signals through the systems. In such cases, it is necessary to provide the conditions for accurate processing of the forcing stimuli.

To establish the limiting capabilities of this class of automatic control system, it is advantageous to investigate the system's behavior when acted upon by sinusoidally varying forcing signals. Other stimuli, varying at finite intervals of time, or with a limited number of nonzero derivatives, determined by the order of the invariant portions of the system, do not give a sufficiently complete representation of the characteristic modes of operations of servo systems and programmed-control systems.

In this connection, the present work investigates the questions related to partial and complete irreproducibility of sinusoidal signals. The partial, or complete, irreproducibility of signals is determined by the presence of limitations placed on the coordinates and derivatives of the system being investigated.

In the present paper, we determine the conditions for partial and complete irreproducibility of a sinusoidal signal and determined the amplitude errors occurring with complete irreproducibility, i.e., we find the limiting frequency characteristics of these systems.

By the limiting frequency characteristics we mean the dependence of the maximum amplitude of the controlled quantity on the frequency, obtained by taking into account the limitations imposed on the system.

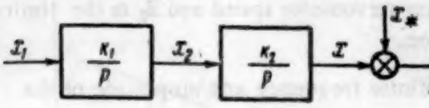


Fig. 1

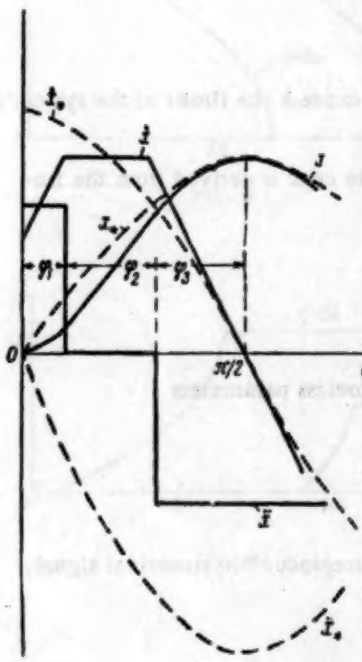


Fig. 2. Reproduction of a sinusoidal signal in the case of limitations on the first and second derivatives, for the optimal control law.

It is obvious with this that the controlled quantity best reproduces the law of variation of the forcing stimulus if the phase shift between them is zero.

As was shown in [1], the problem of investigation optimal processes of reproducing sinusoidal signals depends on whether the forcing stimulus is a reproducible or irreproducible function. In the first case, an optimal reproduction process is meaningful only for the period when the noncorrespondence between the initial system state and the given law of its motion is being removed. After removal of this noncorrespondence, the system motion can coincide completely with the given law of motion.

The results of considering such questions for certain types of reproducible functions were given in [2]. The method applied in [2] may also be used for determining the optimal transient responses in the case when the forcing stimulus varies sinusoidally.

For all the cases to be considered below, we shall assume that the forcing stimulus for the controlled quantity varies sinsoidally:

$$x_*(t) = A \sin \omega t.$$

Following [1], we shall require that the controlled quantity shall best reproduce the variations of the forcing stimulus, in the sense of the closest approximation at the points corresponding to the passage of the function through its peak value.

Starting from this requirement, one finds, by the method presented in [1], the conditions for repducibility of sinusoidal signals without amplitude error, one determines the manner in which the angles of forcing cutoff depend on the system parameters, as well as on the frequency and amplitude of the forcing stimulus. In addition, one finds the dependence of the amplitude error on the characteristic parameters of the system and of the forcing function.

1. Reproduction of Sinusoidal Signals by Astatic Objects Controlled by Integrating Servomotors

The block schematic of the invariant portion of the system considered is shown in Fig. 1. Here, x_1 is the control signal, determining the speed of the servomotor action, x_2 is the path of the controlling organ, x is the controlled quantity, x^* is the forcing function, y is the error signal and $p \equiv d/dt$.

We shall investigate the case when both the coordinate x_1 and the coordinate x_2 are limited, which corresponds to a limitation on the second and first derivatives of the controlled coordinate x^*, i.e., we assume that

$$|\ddot{x}| \leqslant M_1 \text{ and } |\dot{x}| \leqslant M_2$$

where

$$M_1 = k_1 k_2 \bar{x}_1, \quad M_2 = k_2 \bar{x}_2,$$

* The case when only the coordinate x_1 is limited was investigated by A. Ya. Lerner in [1].

$\bar{x}_1$ is the limiting value of the control signal, corresponding to the maximum servomotor speed and $\bar{x}_2$ is the limiting value of the regulating organ's travel, measured from its central position.

The irreproducibility of a sinusoidally varying signal occurs for a definite frequency and amplitude of the forcing function. The condition of irreproducibility has the form

$$\frac{A\omega^2}{M_1} > 1.$$

This inequality shows that the second derivative of the forcing function exceeds the limits of the system's capabilities.

The condition for reproducibility of a sinusoidal signal without amplitude error is derived from the isochronous equation, obtained for a sinusoidal stimulus [2], and has the form

$$\frac{A\omega^2}{M_1} < \frac{\pi}{2}\frac{M_2}{M_1}\omega - \frac{M_2^2}{2M_1^2}\omega^2.$$

By joining these two conditions into one, and by introducing the dimensionless parameters

$$r_1 = \frac{M_2}{M_1}\omega \text{ and } r_2 = \frac{A\omega^2}{M_1},$$

we can obtain the condition for reproducibility without amplitude error of an irreproducible sinusoidal signal, i.e., the condition for partial irreproducibility:

$$1 < r_2 < \frac{\pi}{2} r_1 - \frac{1}{2} r_1^2.$$

If the amplitude and frequency of the signal satisfy these inequalities, then the optimal law for the variation of the controlled quantity will have the form shown in Fig. 2. The optimal control law is characterized by the values of the so-called angles of forcing cutoff, φ_1, φ_2, and φ_3, which are defined by the expressions

$$\varphi_1 = \sqrt{\pi r_1 - r_1^2 - 2r_2}, \quad \varphi_2 = \frac{\pi}{2} - r_1 - \sqrt{\pi r_1 - r_1^2 - 2r_2}, \quad \varphi_3 = r_1.$$

These expressions are obtained by starting from the equations defining the lengths of the switching intervals corresponding to the optimal process [2]. With this, the optimal law is provided by a special controlling portion which closes the system (not shown on Fig. 1). The dependence of the angles of forcing cutoff on the coefficients of derivative excess, denoted previously by r_1 and r_2, is shown in Figs. 3 and 4.

The maximum possible amplitude of the controlled quantity $\underline{x}$ is determined from the condition $\varphi_1 = 0$, and equals

$$A_x \max = \frac{\pi}{2}\frac{M_2}{\omega} - \frac{M_2^2}{2M_1} = \frac{A}{2}\frac{r_1}{r_2}(\pi - r_1).$$

In this case

$$\varphi_2 = \frac{\pi}{2} - r_1, \quad \varphi_3 = r_1.$$

It is shown in [1] that if the amplitude and frequency of the forcing function satisfy a certain inequality then a limited system can not reproduce the forcing function even during the individual portions of the period immediately adjacent to the moment when the forcing function passes through its peak value, and amplitude errors are unavoidable.

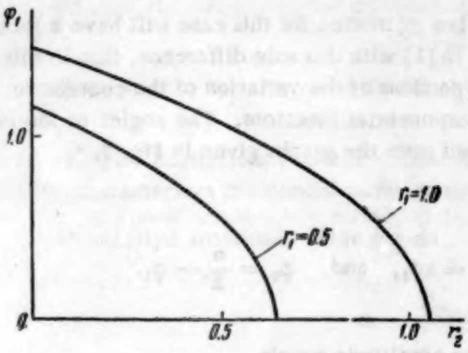


Fig. 3

In the case considered by us, this inequality has the form

$$A > \frac{\pi}{2} = \frac{M_2}{\omega} - \frac{M_2^2}{2M_1}.$$

which corresponds to the mode of complete irreproducibility. With this, the relative amplitude of the reproduced signal equals

$$\alpha = \frac{A_{x\max}}{A} = \frac{1}{2}\,\frac{r_1}{r_2}(\pi - r_1).$$

The functions $\alpha = f(r_1, r_2)$ are shown in Fig. 5. These functions allow one to estimate the limiting capabilities of the system, in the sense of processing sinusoidal signals. By means of system limitations and given frequency and amplitude of the forcing function, the minimum possible amplitude error in the system being considered. Obviously, this error is zero for $\alpha = 1$.

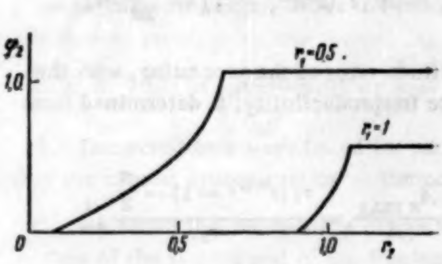


Fig. 4

2. Reproduction of Sinusoidal Signals by a Static Object Controlled by Integrating Servomotors

The block schematic of the invariant portion of the system under consideration is shown in Fig. 6. As in the case of the astatic object, it is of interest to consider two modes of operation: a) when only the coordinate x_1 is limited and b) when the coordinates x_1 and x_2 are limited.

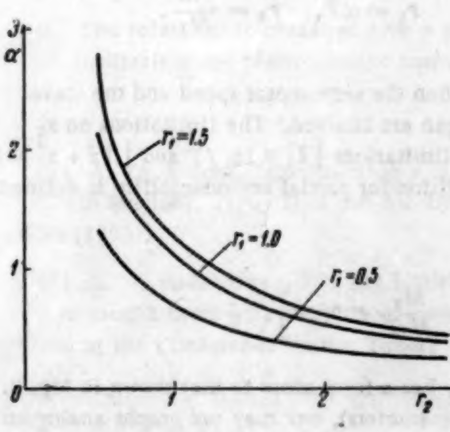


Fig. 5

In case "a", it is the second derivative of the controlled quantity $\underline{x}$ which is limited, and in case "b," a linear combination of derivatives and the coordinate of the controlled quantity $\underline{x}$.

a) The case when only the servomotor speed is limited. The condition for the irreproducibility of the signal is related to the limitation on the coordinate x_1, and leads to the inequality $r^2 > 1$. The condition for the reproducibility of a sinusoidal signal without amplitude error is found analogously to the way given in section 1 above, and has the form

$$r_2 < r_1\left(e^{\pi/2r_1} - 1\right) - \frac{\pi}{2} r_1.$$

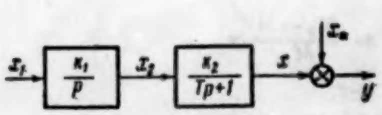


Fig. 6

Then, the condition for partial reproducibility is determined by the inequalities

$$1 < r_2 < r_1\left(e^{\pi/2r_1} - 1\right) - \frac{\pi}{2} r_1.$$

where

$$r_1 = T\omega, \quad r_2 = \frac{TA\omega^2}{M_1}.$$

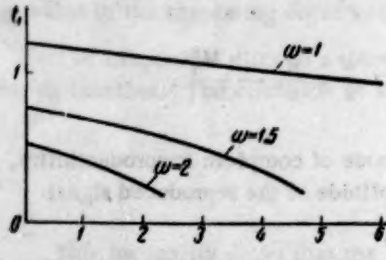


Fig. 7. Dependence of the forcing cutoff time, t_1, on the amplitude, A, of the forcing function.

The optimal law of motion for this case will have a form close to that shown in [1] with this sole difference, that in this case the individual portions of the variation of the coordinate $\underline{x}$ are expressed by exponential functions. The angles of forcing cutoff are determined from the graphs given in Fig. 7. *

With this

$$\varphi_1 = \omega t_1, \quad \text{and} \quad \varphi_2 = \frac{\pi}{2} - \varphi_1.$$

The maximum amplitude equals

$$A_{x\max} = M_1 T\left(e^{\pi/2\omega T} - 1\right) - \frac{\pi}{2\omega} M_1.$$

Then, the amplitude error of the processing, with the condition for complete irreproducibility, is determined from the expression

$$\alpha = \frac{A_x \max}{A} = \frac{r_1\left(e^{\pi/2r_1} - 1\right) - \frac{\pi}{2}}{r_2},$$

where

$$r_1 = \omega T, \quad r_2 = \frac{A\omega}{M_1}.$$

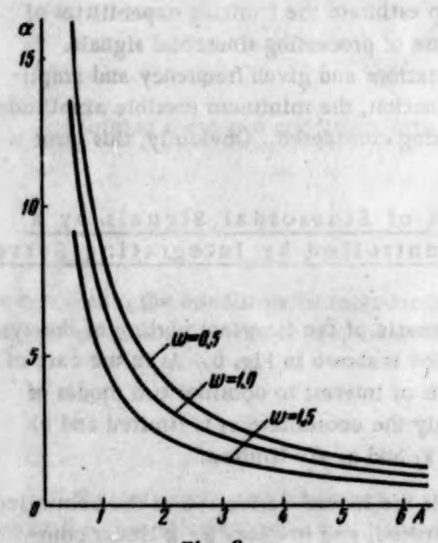


Fig. 8.

b) The case when the servomotor speed and the travel of the controlling organ are limited.* The limitations on x_1 and x_2 reduce to the limitations $|\ddot{x}| \le M_1/T$ and $|T\dot{x} + x| \le M_2$. In this case, the condition for partial reproducibility is defined by the inequality

$$1 < \frac{TA\omega^2}{M_1} < \frac{M_2}{M_1} T\omega^2 + T^2\omega^2 \ln\left[1 - \frac{M_2}{M_1 T} e^{-\pi/2\omega T}\right].$$

With this, the optimal law of variation of the controlled quantity has a form close to that shown in Fig. 2. For finding the angles of forcing cutoff in this case (for given system parameters), one may use graphs analogous to those of Fig. 7.

Then,

$$\varphi_1 = \omega t_1, \quad \varphi_2 = \frac{\pi}{2} - \frac{M_2}{M_1}\omega + \frac{A\omega}{M_1} - \varphi_1, \quad \varphi_3 = \frac{M_2 - A}{M_1}\omega.$$

The maximum possible value of amplitude of the controlled quantity equals

$$A_x \max = M_2 + M_1 T \ln\left[1 - \frac{M_2}{M_1 T} e^{-\pi/2\omega T}\right].$$

* The graphs are constructed for the system in which $M_1 = 6$, $T = 1$.

For

$$A > M_2 + M_1 T \ln\left[1 - \frac{M_2}{M_1 T} e^{-\pi/2\omega T}\right]$$

the system with limitations imposed cannot reproduce the forcing function without amplitude errors. This inequality characterizes the condition for complete irreproducibility.

The relative amplitude value equals

$$\alpha = \frac{A_{x\max}}{A} = \frac{r_4 + \ln\left[1 - r_4 e^{-\pi/2r_1}\right]}{r_5},$$

where $r_1 = \omega T$, $r_4 = M_2/M_1T$ and $r_5 = A/M_1T$.

As an example, Fig. 8 shows the function $\alpha = f(A, \omega)$ for the systems with the parameters: $M_1 = 6$, $M_2 = 9$ and $T = 1$.

SUMMARY

1. The conditions were found for partial, and complete, irreproducibility of sinusoidal signals, in the sense of the closest approximation at the points corresponding to the peak value of the forcing function.

2. We found the dependence of the forcing cutoff angles, i.e., the control laws, on the characteristic parameters of the system and of the forcing function.

3. We obtained expressions showing the dependence of the amplitude errors on the system parameters for complete irreproducibility of the signal.

4. The relationship obtained give a picture of the limiting frequency properties of the systems considered, in which limitations are placed on the controlled quantity and its derivatives.

LITERATURE CITED

[1] A. Ya. Lerner, "Constructing a fast-acting SAR with limitations on the coordinate of the controlled object," [in Russian]. Trudy II of the All-Union Conference on Automatic Control Theory, vol. II. Izdatelstvo AN SSSR (1955).

[2] G. A. Nadzhafova, "On the limiting dynamic indicators of standard objects, controlled by astatic executive mechanisms with limitations on the speed and travel of the controlling organ," [in Russian]. Paper delivered at the Conference on the Theory and Application of Discrete Automatic Systems. IAT AN SSSR (1958).

Received October 10, 1958

ON THE ANALYSIS AND SYNTHESIS OF CERTAIN ELECTRICAL CIRCUITS BY MEANS OF SPECIAL LOGICAL OPERATORS

A. D. Talantsev

(Moscow)

Special logical operators are defined which describe the change in Boolean functions as their arguments change. The properties of these operators are investigated, and transformation formulae are derived which, in many cases, allow simplification of the circuit to be synthesized. As an example, we consider the transformation of a potential-pulse circuit used in a digital control system for a milling machine.

INTRODUCTION

Logical-algebraic methods are used in the investigation of various types of circuits for transforming information. Thus, in 1938 V. I. Shestakov [1] and C. Shannon [2] simultaneously developed the interpretation, in terms of relay-contact circuits, of the combinatorial propositional and one-argument predicate calculus (Boolean algebra). From that thime on, the application of mathematical logic was intensively developed. A new impulsion to its development was provided by the appearance of high-speed computers for arithmetic, and nonarithmetic, tasks. Electronic logical circuits are widely used in these machines. Adequate methods were developed for the analysis and synthesis of potential and pulse circuits (Cf., for example [3-5]). The so-called potential-pulse circuits, in which there is simultaneously a potential mode and a pulse mode came into widespread use. In these circuits, the generation of a pulse is usually related to the state of some element. An example might be the formation of a transfer pulse at the output of a flip-flop.

The circuit elements are described by logical variables (one-argument predicates), in which time plays the role of the single variable. We note, however, that with existing methods, the manner in which the logical variables are given in time is not brought into question. Potential and pulse modes are described formally in the identical way. Such a description only establishes conventionally the fact of a change in a variable's value after substitution of the time, $\underline{t}$, is made. This leads to a limitation of the analysis of such circuits, since it is impossible to correlate this concept of "change in value" with several logical variables, or with functions of them.

In this paper, we attempt to remove this limitation. To describe the change of state of elements, we introduce special logical operators on logical variables. These operators are two-argument predicates of predicates, in which one argument is a one-argument predicate and the other is an individual variable. *

1. Potential and Pulse Logical Variables

We consider an electrical circuit in which information is represented by binary digits (bits). Each of the

* The basic content of this paper was given in seminars on the technical applications of mathematical logic, at the Moscow State University on October 2, 1958 and January 16, 1959.

input and ouput buses of such a circuit can be found in one of two stages, distinguished, for example, by voltage level, polarity, etc. One of these states is assigned the sign "1" ("one state") and the other is assigned the sign "0" ("zero state"). The signs 1 and 0 are the values of the logical variable $Z(t)$. The values of the variable $Z(t)$ may be given over intervals of time, or at discrete moments of time.

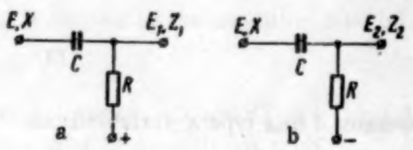


Fig. 1

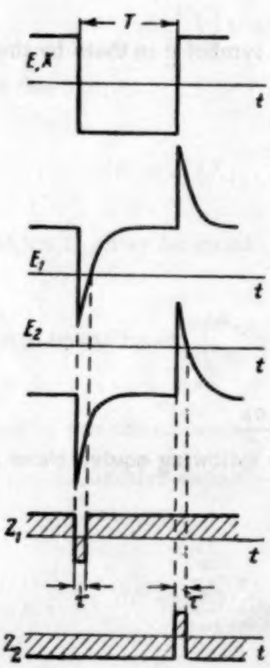


Fig. 2

We shall assert that the unit (respectively, zero) value of $Z(t)$ is given on the interval (t_k, t_l) if: 1) for all t satisfying the condition $t_k < t < t_l$, $Z(t) = 1$ (respectively, $Z(t) = 0$; 2) $Z(t_k) = 0$, $Z(t_l) = 0$ (respectively, $Z(t_k = 1)$, $Z(t_l) = 1$).

We shall assert that the unit value of $Z(t)$ at time t is given discretely if: 1) $Z(t) = 1$; 2) there exists an interval of time ϵ such that, for all t' satisfying the condition $t - \epsilon \leq t' < t$ and for all t'' satisfying the condition $t < t'' \leq t + \epsilon$, $Z(t') = 0$ and $Z(t'') = 0$.

The variable $Z(t)$ will be called a potential variable, and will be denoted by $X(t)$, if its values, both ones and zeroes, are given only on intervals, and these intervals are not less than some given time interval T.

The variable $Z(t)$ will be called a pulse variable, and will be denoted by $Y(t)$, if: 1) its unit values are given either only discretely or only on intervals not greater than the time interval $\tau << T$; 2) its zero values are given only intervals not less than the time interval $T - \tau$.

The time intervals T and τ characterize, respectively, the potential and the pulse modes of circuit operation.

2. Transition Operators on Logical Variables

We shall say that, at the moment of time t, the unit value of the variable $Z(t)$ changes to the zero value if there exists a time interval $\delta > \tau$ such that $Z(t') = 1$ and $Z(t'') = 0$ for all moments of time t' and t'' such that $t - \delta \leq t' < t$, and $t < t'' \leq t + \delta$.

Analogously, we shall say that, at the moment of time t, the zero value of $Z(t)$ changes to the unit value if there exists a time interval $\delta > \tau$ such that $Z(t') = 0$ and $Z(t'') = 1$ for all moments of time t' and t'' such that $t - \delta \leq t' < t$ and $t < t'' \leq t + \delta$.

The presence (respectively, absence) at time t of a transition from the unit value of $Z(t)$ to the zero value will be described by the equation $dZ(t) = 1$ (respectively, $dZ(t) = 0$). Then,the presence (absence) at time t of a transition from the zero value of $Z(t)$ to the unit value will be expressed by the equation $d\overline{Z}(t) = 1$ (respectively, $d\overline{Z}(t) = 0$).

We shall say that, at time t, the variable $Z(t)$ changes its value if the equality $DZ(t) = 1$ holds, where

$$DZ(t) = dZ(t) \vee d\overline{Z}(t). \tag{1}$$

We shall agree to consider the time t to be the same for all $Z(t)$ and, further, we shall henceforth omit the letter "t" from the formulae. The symbols d and D we call transition operators on logical variables.

We note certain properties of the operators d and D.

1. Application of the operators d and D to a type X variable leads to a type Y variable.

2. **Application of operator $\underline{d}$ to a type Y variable leads to a variable which is identically zero:**

$$dY = 0. \tag{2}$$

3. **Application of operator $\underline{d}$ to a constant leads to an identically zero variable:**

$$d1 = 0, \quad d0 = 0. \tag{3}$$

It follows from statements 1, 2 and 3 that repeated application of operator $\underline{d}$ to a type X variable leads to an identically zero variable:

$$d \ldots dX = 0 \tag{4}$$

Results (2), (3) and (4) are extended to operator D by simply replacing the symbol $\underline{d}$ in them by the symbol D. The operators $\underline{d}$ and D will henceforth be applied only to type X variables.

4. **We cannot simultaneously have $dX = 1$ and $d\overline{X} = 1$, i.e.,**

$$dX \,\&\, d\overline{X} = 0. \tag{5}$$

5. **The following equivalence holds:**

$$D\overline{X} = DX. \tag{6}$$

3. Application of Transition Operators to Boolean Functions

We first apply the operators $\underline{d}$ and D to certain elementary functions. The following equivalences are easily established: *

$$d(X_1 \vee X_2) = \overline{X}_1 dX_2 \vee \overline{X}_2 dX_1 \vee dX_1 dX_2, \tag{7}$$

$$d(X_1 X_2) = X_1 dX_2 \vee X_2 dX_1 \vee dX_1 dX_2, \tag{8}$$

$$D(X_1 \vee X_2) = \overline{X}_1 DX_2 \vee \overline{X}_2 DX_1 \vee dX_1 dX_2 \vee d\overline{X}_1 d\overline{X}_2, \tag{9}$$

$$D(X_1 X_2) = X_1 DX_2 \vee X_2 DX_1 \vee dX_1 dX_2 \vee d\overline{X}_1 d\overline{X}_2. \tag{10}$$

Consider, for example, the left member of (7). A unit value of $X_1 \quad X_2$ leads to a zero in one of three cases: 1) when, for $X_1 = 0$, a unit value of X_2 changes to zero; 2) when, for $X_2 = 0$, a unit value of X_1 changes to zero and 3) when X_1 and X_2 simultaneously change from value 1 to value 0. These cases, in the form of a disjunction of conjunctions, are enumerated in the right member of (7). The right members of (7), (8), (9) and (10) are called disjunctive expansions of the corresponding operators $\underline{d}$ and D as applied to disjunctions and conjunctions.

We now apply the operators $\underline{d}$ and D to an arbitrary Boolean function $F(X_1, X_2, \ldots, X_n)$. We introduce the notation, $\sum_{i=1}^{n} X_i = X_1 \vee X_2 \vee \ldots \vee X_n$, $\prod_{i=1}^{n} X_i = X_1 \,\&\, X_2 \,\&\, \ldots \,\&\, X_n$, $X^0 = \overline{X}$, $X^1 = X$. The

*** The sign & will be frequently omitted in what follows.**

disjunctive expansion of dF is a disjunction of conjunctions, each of which contains one of the terms $dX_{i_1}^{p_1} dX_{i_2}^{p_2} \ldots dX_{i_k}^{p_k}$ ($i_s = 1, 2, \ldots, n$; $p_s = 0, 1$; $s = 1, 2, \ldots, k$; $k = 1, 2, \ldots, n$) and the Boolean function $Q_{p_1, p_2, \ldots, p_k}^{i_1, i_2, \ldots, i_k}$ of variables from the set $X_1, X_2, \ldots, X_n$ which do not follow the d in the conjunction considered. The function Q is defined by the condition that the equality dF = 1 exist with the equality $\prod_{s=1}^{k} dX_{i_s}^{p_s} = 1$. For example, let

dF = 1 for $dX_i = 1$ and $d\overline{X}_j = 1$ ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$; $i \neq j$). For this, it is necessary and sufficient that the following conditions hold:

$$F(X_1, \ldots, X_i, \ldots, X_j, \ldots, X_n) = 1 \quad \text{for} \quad X_i = 1, \quad X_j = 0,$$
$$F(X_1, \ldots, X_i, \ldots X_j, \ldots, X_n) = 0 \quad \text{for} \quad X_i = 0, \quad X_j = 1,$$

i.e., in this case

$$dF = F(X_1, \ldots, 1, \ldots, 0, \ldots, X_n)\, \overline{F}(X_1, \ldots, 0, \ldots, 1, \ldots, X_n)\, dX_i\, d\overline{X}_j,$$

from which Q_{10} may be found

We denote by $F\begin{pmatrix} X_{i_1}, \ldots, X_{i_k} \\ P_1, \ldots, P_k \end{pmatrix}$ the function $F(X_1, X_2, \ldots, X_n)$ in which the variables $X_{i_1}, \ldots, X_{i_k}$ are replaced by the corresponding arbitrary Boolean functions $P_1, \ldots P_k$.

The disjunctive expansion of dF has the form

$$dF = \sum_{k=1}^{n} \sum_{p_1=0}^{1} \cdots \sum_{p_k=0}^{1} \sum_{i_1, \ldots, i_k}^{n} F\begin{pmatrix} X_{i_1}, \ldots, X_{i_k} \\ p_1, \ldots, p_k \end{pmatrix} \overline{F}\begin{pmatrix} X_{i_1}, \ldots, X_{i_k} \\ \bar{p}_1, \ldots, \bar{p}_k \end{pmatrix} \prod_{s=1}^{k} dX_{i_s}^{p_s}. \tag{11}$$

Here, $\sum_{i_1, \ldots, i_k}^{n}$ denotes the disjunction over all possible sets of the indexes $1, 2, \ldots, n$ taken $\underline{k}$ at a time.

The result is analogous for DF. It may be obtained from the equation

$$DF = dF \vee d\overline{F}.$$

The number of different terms of the form $\prod_{s=1}^{k} dX_{i_s}^{p_s}$ ($i_s = 1, 2, \ldots, n$; $p_s = 0, 1$; $k = 1, 2, \ldots, n$) equals $3^n - 1$.

We note, however, that for any Boolean function $F(X_1, X_2, \ldots, X_n)$, the number of disjunctive terms in the expansion of dF does not exceed $3^n - 2^{n-1} - 1$ since, from the group of terms in the right member of (11) which correspond to k = n, we eliminate that half of the terms which all equal zero.

At any moment of time, let no more than one of the variables from the set $X_1, X_2, \ldots, X_n$ change its value. We call this limitation the "condition of noncoincident transitions." For this case, which is important in the applications, the formulae for dF and DF are simplified:

$$dF = \sum_{p=0}^{1} \sum_{i=1}^{n} F \begin{pmatrix} X_i \\ p \end{pmatrix} \overline{F} \begin{pmatrix} X_i \\ \overline{p} \end{pmatrix} dX_i^p, \tag{12}$$

$$DF = \sum_{i=1}^{n} \left[F \begin{pmatrix} X_i \\ 1 \end{pmatrix} \overline{F} \begin{pmatrix} X_i \\ 0 \end{pmatrix} \vee \overline{F} \begin{pmatrix} X_i \\ 1 \end{pmatrix} F \begin{pmatrix} X_i \\ 0 \end{pmatrix} \right] DX_i. \tag{13}$$

Note. It is interesting to note the analogy between expression (8) and the well-known formula of mathematical analysis, d(uv) = udv + vdu + dudv. The analogy is based on the correspondences of conjunction with arithmetic multiplication and of disjunction with addition. The term dudv, a higher-order infinitesimal, corresponds to the term dX_1dX_2 which equals zero if the condition of noncoincident transitions holds. However, the analogy immediately breaks down when one turns to expressions (7), (9) and (10), which contain the operation of logical negation. Nonetheless, the operator $\underline{d}$ might be called the "differential" of a logical function.

4. On the Analysis of Homogeneous Potential-Pulse Circuits

In this paper, only that type of circuit will be considered in which all input buses are described by potential variables and all output buses by pulse variables. For brevity, such circuits will sometimes be referred to below simply as "circuits."

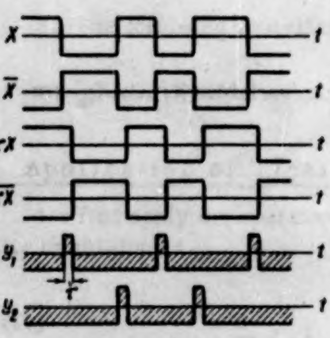


Fig. 3

We consider a homogeneous potential-pulse circuit with $\underline{n}$ input buses and $\underline{r}$ outputs. Let the $\underline{n}$ input buses be described by the variables $X_1, X_2, \ldots, X_n$ and the $\underline{r}$ output buses by the variables $Y_1, Y_2, \ldots, Y_r$. We assume that such a circuit possesses the following properties.

If, at time $\underline{t}$, the variable Y_j ($j = 1,2,\ldots,r$) equals unity, then at least one of the variables in the set $X_1, X_2, \ldots, X_n$ changed value at one of the moments, t', which satisfy the condition $-\infty < t' \leq t$.

The circuit being considered is a scheme for transforming the information given by the set $X_1, X_2, \ldots, X_n$, into information defined by the set $Y_1, Y_2, \ldots, Y_r$. Such a circuit must certainly contain elements which "transform" the potential variables $X_1, X_2, \ldots, X_n$ into the pulse variables $Y_1, Y_2, \ldots, Y_r$. Let us consider examples of such "transformers."

Figures 1a and 1b show the simplest RC differentiating circuits. To differentiate the bus states, we denote a positive voltage by 1 and a negative voltage by 0. The plus and minus signs on the figure denote the corresponding time-invariant positive and negative voltages.

Figure 2 shows the curves for the voltages E, E_1 and E_2, and the corresponding logical variables X, Z_1 and Z_2. In the first case, for $Z_1 = 0$ at time $\underline{t}$, we have dx = 1 for any time t' which satisfies the condition $t - \tau \leq t' \leq t$. With a properly chosen time constant, RC, the magnitude of τ may be made very small, so we therefore assume that $Z_1 = d\overline{X}$ and $Z_2 = d\overline{X}$.

The transformation of a potential variable into a pulse variable can also be implemented by means of the operations of Boolean algebra and the time delay operator.

We shall say that the function $\tau X(t)$ is the variable X(t) delayed by the time interval τ if

$$\tau X(t) = X(t - \tau). \tag{14}$$

Here, the symbol τ before X(t) plays the role of the time delay operator on the variable X(t).

We now define two pulse variables, Y_1 and Y_2, as follows:

$$Y_1 = \overline{X} \,\&\, \tau X, \tag{15}$$
$$Y_2 = X \,\&\, \overline{\tau X}. \tag{16}$$

It is easily understood from Fig. 3 that, if $Y_1 = 1$ (respectively, $Y_1 = 1$) at time $\underline{t}$, then $dX(t) = 1$ (respectively, $d\overline{X}(t) = 1$) for any time t' which satisfies the condition $t - \tau \leq t' \leq t$. By choosing the quantity τ sufficiently small (but not equal to zero), we may assume that $Y_1 = dX$ and $Y_2 = d\overline{X}$.

This method of transformation, called "digital differentiation," is used to obtain carry pulses in counting circuits constructed of dynamic elements [6].

We now make the logical structure of homogeneous potential-pulse circuits more precise. Not all changes in value of some variable X_i, generally speaking, lead to an equation $Y_j = 1$. In the general case, $Y_j = 1$ at time $\underline{t}$ when there are present conditions defined by the values of several elements of the set $X_1, X_2, \ldots, X_n$ at time t', such that $-\infty < t' \leq t$. The variable Y_j may, therefore, be given in the form of a disjunctive expansion.

$$Y_j = \sum_{k=1}^{n} \sum_{p_1=0}^{1} \ldots \sum_{p_k=0}^{1} \sum_{i_1,\ldots,i_k}^{n} R_{\substack{p_1,\ldots,p_k \\ i_1,\ldots,i_k}} \prod_{s=1}^{k} dX_{i_s}^{p_s}, \tag{17}$$

where $R_{\substack{p_1,\ldots,p_k \\ i_1,\ldots,i_k}}$ is some Boolean function of variables from the set $X_1, X_2, \ldots, X_n$ considered for $t' \leq t$.

The right member of (17) is called the potential-pulse form. Here, two cases should be mentioned: 1) the function $R_{\substack{p_1,\ldots,p_k \\ i_1,\ldots,i_k}}$ is considered for $t' = t$; in this case it may be given only by those variables of the set $X_1, X_2, \ldots, X_n$ which are not subject to the operation $\underline{d}$ in the terms with the subscript $\substack{p_1,\ldots,p_k \\ i_1,\ldots,i_k}$ 2) the function $R_{\substack{p_1,\ldots,p_k \\ i_1,\ldots,i_k}}$ is considered for the case when $t' < t$; in this case it may be given by all the variables $X_1, X_2, \ldots, X_n$, except for those X_i for which $dX_i^p(t') = 1$. In the first case we can speak of a circuit of the normal type, and in the second of a circuit with delays. In what follows we shall be speaking only of normal-type circuits.

5. On Integrating Potential-Pulse Forms

Let a circuit described by form (17) be given. The question naturally arises: can a function F be found such that $Y_j = dF$? The finding of such an "original" function would make it possible to simplify significantly form (17) as a result of reducing the number of $\underline{d}$ signs and conjunctions entering into it.

If there exists a function F such that $Y_j = dF$, then form (17) is called an integrable potential-pulse form. The finding of the original function F is called integrating form (17).

If form (17) is integrable, then the following identity must hold for the original function F:

$$F\begin{pmatrix} X_{i_1}, \ldots, X_{i_k} \\ p_1, \ldots, p_k \end{pmatrix} \overline{F}\begin{pmatrix} X_{i_1}, \ldots, X_{i_k} \\ \overline{p}_1, \ldots, \overline{p}_k \end{pmatrix} \equiv R_{\substack{p_1,\ldots,p_k \\ i_1,\ldots,i_k}}. \tag{18}$$

Integration of form (17) can be carried out in the following way. A system of identities of the type shown in (18) is set up. From this system are found the values of the coefficients of the complete disjunctive normal form, $F(p_1, p_2, \ldots, p_n)$. A set of 2^n of such coefficients, is chosen, defing the function F. One then determines whether or not the values of the remaining coefficients of $F(p_1, p_2, \ldots, p_n)$, determined from the system in (18), are contradicted by the values of the corresponding coefficients in the set cited. If there is no contradiction, then this set defines the function F such that $Y_j = dF$.

For example, let the following form be given:

$$
\begin{aligned}
Y = \overset{1}{\overline{X_2X_3}}\, dX_1 \vee \overline{X}_1\overset{2}{X_3}\, dX_2 \vee \overline{X}_1\overset{3}{X_2}\, dX_3 \vee d\overset{4}{X_1}\, dX_2 \vee \\
\vee\, d\overset{5}{X_1} dX_3 \vee \overline{X}_1 d\overset{6}{X_2} dX_3 \vee \overline{X}_3 d\overset{7}{X_1} d\overline{X}_2 \vee \overline{X}_2\, d\overset{8}{X_1} d\overline{X}_3 \vee \\
\vee\, dX_1\, \overset{9}{d}X_2\, dX_3 \vee dX_1\, d\overset{10}{\overline{X}_2}\, dX_3 \vee dX_1\, \overset{11}{d}X_2\, d\overline{X}_3.
\end{aligned}
\qquad (19)
$$

We set up the two systems of identities:

$$
\begin{aligned}
& R_{\substack{1\\1}} = F(1X_2X_3)\,\overline{F}(0\,X_2X_3) \equiv \overline{X_2X_3}, \\
& R_{\substack{1\\2}} = F(X_1 1\, X_3)\,\overline{F}(X_1 0 X_3) \equiv \overline{X}_1 X_3; \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& R_{\substack{11\\12}} = F(11X_3)\,\overline{F}(0\ 0\ X_3) \equiv 1, \\
& R_{\substack{11\\13}} = F(1\,X_2\,1)\,\overline{F}(0\,X_2\,0) \equiv 1, \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& R_{\substack{111\\123}} = F(111)\,\overline{F}(0\ 0\ 0) = 1,
\end{aligned}
\qquad (20)
$$

$$
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
$$

$$
R_{\substack{0\\1}} \equiv 0,\ R_{\substack{0\\2}} \equiv 0, \ldots,\ R_{\substack{01\\12}} \equiv 0, \ldots \qquad (21)
$$

We find from system (20) that

$$
\begin{aligned}
& F(000) = 0,\ F(001) = 0,\ F(010) = 0,\ F(011) = 1, \\
& F(100) = 1,\ F(101) = 1,\ F(110) = 1,\ F(111) = 1.
\end{aligned}
\qquad (22)
$$

We then find the values of $F(p_1p_2p_3)$ obtained from this system, and also the values of the remaining coefficients from system (20) do not contradict the values of the corresponding coefficients from the set in (22). This set defines the function $F = X_1 \quad X_2X_3$, i.e.,

$$
Y = d(X_1 \vee X_2X_3).
$$

If, in this example, the variables X_1, X_2 and X_3 satisfy the condition of noncoincidence of transitions, then the order of choosing the set of coefficients of $F(p_1p_2p_3)$ is somewhat different. In this case, terms 4 through 11 in (19) are lacking, and the system of identities in (21) is also used for choosing the set in (22).

When there are contradictory values of the coefficients of $F(p_1, p_2, \ldots, p_n)$ for one and the same set, $p_1, p_2, \ldots, p_n$, the terms of formula (17), by means of redundant combinations, are divided into groups, to each of which the method stated is applied. Thus, we arrive at the representation

$$
Y_j = \sum_{i=1}^{m} dF_i. \qquad (23)
$$

If the condition of noncoincidence of transitions holds for the variables $X_1, X_2, \ldots, X_n$, the potential-pulse form (17) may be given in the form

$$Y_j = \sum_{s=1}^{m} dF_s(X_1, X_2, \ldots, X_n) \sum_{i=1}^{n} \sum_{p=0}^{1} q_{sip}\, dX_i^p. \tag{24}$$

For this, it is necessary to find $\underline{m}$ functions F_s and 2mn coefficients $q_{sip} = 0, 1$.

6. The Transformation of One Logical Circuit

For measuring the spatial translations in systems for the digital control of actual objects, the so-called "quantized scale" is employed. Figure 4 shows the scale with separated point intervals which is frequently used in digital control systems for milling machines [7]. The scale is divided into equal intervals ("quanta") or white ("transparent") and black ("opaque") points. The scale is read by two photoelements, b_1 and b_2, displaced from each other by half a scale quantum. The light flux from the light source placed on the other side of the scale (not shown on the figure) excites, or does not excite, the photoelements. The photoelement b_i ($i = 1,2$) is characterized by the variable X_i ($i = 1,2$), which equals one if a black point of the scale is opposite b_i, and equals zero if a white point is opposite b_i.

To control a movable object, it is necessary to define the direction of motion of the scale. For this purpose, we introduce two logical functions, φ and ψ, which can take the value of one only when the scale is moving, respectively, from right to left and from left to right.

It follows, from a direct analysis of scale motion according to Fig. 4, that

$$\varphi = X_1^1 dX_2 \vee X_2^2 d\overline{X}_1 \vee \overline{X}_2^3 dX_1 \vee \overline{X}_1^4 d\overline{X}_2, \tag{25}$$

$$\psi = X_2 dX_1 \vee X_1 d\overline{X}_2 \vee \overline{X}_1 dX_2 \vee \overline{X}_2 d\overline{X}_1. \tag{26}$$

The functions φ and ψ are realized by a homogeneous potential-pulse circuit with two input and two output buses. The input buses are described by the functions X_1 and X_2 and the output buses by the functions φ and ψ.

Fig. 4

Let us try to integrate the potential-pulse forms (25) and (26). We shall begin with the integration of form (25). We first set up the system of identities of the type shown in (20). We obtain contradictory values of the coefficients of $F(p_1p_2)$, from which we conclude that form (25) is not integrable. We then divide the terms of form (25) into groups in the following way: term 1 is joined with term 4, and term 2 is joined with term 3. We now integrate each group separately For the group of terms 1 and 4 we get the function F_1.

We set up the system of the type in (20):

$$\underset{2}{R_1} = F_1(X_1 1)\overline{F}(X_1 0) \equiv X_1,$$
$$\underset{2}{R_0} = F_1(X_1 0)\overline{F}(X_1 1) \equiv \overline{X}_1. \tag{27}$$

From this we find that

$$F(11) = 1,\ F(00) = 1,\ F(10) = 0,\ F(01) = 0. \tag{28}$$

We now set up a system of the type shown in (21):

$$\underset{1}{R_1} = F(1X_2)\overline{F}_1(0X_2) \equiv 0,\ \underset{1}{R_0} = F_1(0X_2)\overline{F}_1(1X_2) \equiv 0. \tag{29}$$

From this system we do obtain values of $F_1(p_1p_2)$ which contradict the values of the coefficients from the set in (28). Thus, for example, for F_1 (11) = 1, the value of F_1 (01) also equals unity. Thus, the group of terms 1 and 4 from (25) is not integrable. However, the condition of noncoincidence of transitions holds for the variables X_1 and X_2. Therefore, we use the possibility of presenting form (25) in the form of (24).

The set in (28) defines the function $F_1 = X_1X_2 \quad \overline{X}_1 \overline{X}_2$, for which

$$dF_1 = \overset{1}{X_2} dX_1 \vee \overset{2}{\overline{X}_2} d\overline{X}_1 \vee \overset{3}{X_1} dX_2 \vee \overset{4}{\overline{X}_1} d\overline{X}_2.$$

In this expansion, terms 3 and 4 correspond to the terms 1 and 4 of (25), and terms 1 and 2 are "superfluous." These latter are eliminated by the choice of the coefficients q_{sip}. We let

$$q_{111} = 0, \ q_{110} = 0, \ q_{121} = 1, \ q_{120} = 1.$$

For the group of terms 2 and 3 of form (25), we find that $F_2 = X_1\overline{X}_2 \quad \overline{X}_1X_2$ and

$$q_{211} = 1, \ q_{210} = 1, \ q_{221} = 0, \ q_{220} = 0.$$

We apply analogous transformations to form (26). Finally, the system of functions φ, ψ takes the form:

$$\varphi = d(X_1X_2 \vee \overline{X}_1\overline{X}_2)\, DX_2 \vee d\,\overline{(X_1X_2 \vee \overline{X}_1\overline{X}_2)}\, DX_1, \tag{30}$$

$$\psi = d(X_1X_2 \vee \overline{X}_1\overline{X}_2)\, DX_1 \vee d\,\overline{(X_1X_2 \vee \overline{X}_1\overline{X}_2)}\, DX_2. \tag{31}$$

System (30), (31) requires fewer physical elements for its realization than does the system consisting of (25) and (26). The decrease is due to the decrease in the number of different conjunctions entering into this system: six conjunctions instead of eight.

SUMMARY

From what has been presented above, the following conclusions may be drawn.

1. Transition operators are advantageously used in those cases when it is necessary to describe symbolically the changes in state of some circuit elements occurring under conditions determined by the states of other elements.

2. Circuit analysis by means of transition operators makes it possible to employ new methods of simplifying circuits and, therefore, permits greater efficiency in designing circuits so as to make them approach the ideal.

3. The application of transition operators to the problem of determining the direct of motion of a quantized scale allowed a new circuit to be obtained purely by formal means which turned out to be simpler than the well-known circuits.

The author wishes to express his appreciation to V. A. Trapeznikov, V. I. Shestakov and M. L. Tsetlin for their interest in this work and for their valuable remarks in the discussions of the results.

LITERATURE CITED

[1] V. I. Shestakov. Some Mathematical Methods for Designing and Simplifying Two-Terminal Class A Circuits. [in Russian]. Dissertation for the degree of uchenoi stepeni kand. fiz.-mat. nauk (1938).

[2] C. E. Shannon, A Symbolic Analysis of Relay and Switching Circuits. Transaction of AIEE, 1938, vol. 57, pp. 713-722.

[3] The Synthesis of Electronic Computing and Controlling Systems. [in Russian, translated from English under the direction of V. I. Shestakov]. ILI (1954).

[4] E. C. Nelson. An Algebraic Theory for Use in Digital Computer Design, Trans. IRE, vol. EC-3, 1954, No. 3, p. 12-21.

[5] M. L. Tsetlin, "A matrix method for the analysis and synthesis of electronic-pulse and relay-contact (nonprimitive) circuits," [in Russian]. Dokl. AN SSSR, 117, 6 (1957).

[6] B. Cohen. A Formal Procedure for the Logical Design of an Optimum Binary Counter. Proc. NEC, vol. 10, 1954.

[7] A. D. Talantsev, "Transformation of angle of rotation and linear displacement into digital form," [in Russian]. Automation and Remote Control (USSR) 20, 3 (1959).

Received November 1, 1958

A METHOD FOR COMPUTING THE CHARACTERISTICS OF A DC MOTOR WITH THROTTLE CONTROL

D. A. Alenchikov and V. S. Kulebakin

(Moscow)

A method is developed for calculating the characteristics of dc electric drives with throttle control [1]. The method is based on the volt-ampere characteristics of the saturable core and its load, and also on the use of operating point trajectories, which are constructed graphically from the family of load volt-ampere characteristics.

The purpose of the present paper is to develop a method of computing the characteristics of energy indicators of dc electric drives with throttle control [1]. For definiteness and for great generality, the method of calculation is given for the example of a throttle-controlled motor with series excitation.

Figure 1 gives the block schematic for the control of a series-excited motor supplied from a three-phase ac source via a Larionov rectifier bridge, as well as the circuit for one-phase conversion and the vector diagram of the voltages in the conversion circuit. We may as consider as invariant, in practice, only the ohmic impedance of the conductors in the armature network of the conversion circuit, if we neglect the effect of temperature changes. The impedance of the brush contacts and the rectifier, r_D, depend on the current. The active impedance of the saturable core (SC), which is determined by the ohmic and additional losses in the copper and steel, varies with changing core current and voltage.

The rotational counter emf of the series-excited motor armature depends on the current in the armature network (excitation current) the extent of saturation of the individual portions of the magnetic circuit, the armature reaction and the speed of rotation. In computing the characteristics of a nonreversible electric drive, it is necessary to take the residual magnetism into account.

If selenium rectifiers are used, neglecting of their impedance can lead to large errors, particularly in the region of small values of rotational counter emf. The error becomes particularly large when the selenium rectifiers are air-conditioned to increase the specific load.

The load on the saturable core (the motor, the semiconducting diodes) and the saturable core can be characterized by families of volt-ampere characteristics. The volt-ampere characteristics of saturable cores are usually given in specific units, and are widely used in designing circuits without self-saturation (Cf. for example, [2]).

In this paper, we give the volt-ampere characteristics of a saturable core for the case when the loaded (working) windings are connected by a self-saturating circuit with a current positive feedback gain of unity. The characteristics in specific units are shown in the first quadrant of Fig. 2.

The specific voltage drop across the core, $\underline{u}$, is expressed in terms of the absolute voltage U_{sat} in the following way:

$$u = \frac{C_1 \cdot 10^6}{f S_c w_\sim} U_{sat} = k_1 U_{sat} \tag{1}$$

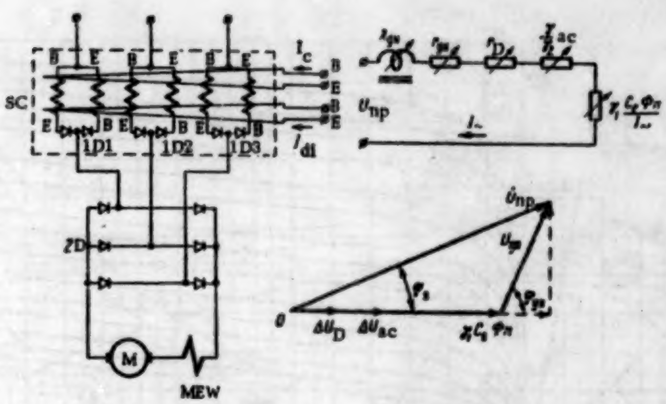


Fig. 1. SC is the three-phase saturable core, 1D1 – 1D3 are feedback diodes, 2D is the three-phase Larionov bridge rectifier, M is the series-excited motor, MEW is the motor's excitation winding, I_c is the saturable core's control current and I_{dl} is the disturbance current of the saturable core. The direction of coil winding is from the beginning (B) to the end (E).

Here, f is the frequency of the voltage applied to the core; S_c is the cross section of the steel magnetic conductor, in cm^2; $w_\sim$ is the number of turns in one loaded winding, C_1 is a coefficient which indicates what portion of the voltage drop across the core comprises the voltage drop across one loaded winding (for the circuit of Fig. 1, $C_1 \approx 1$), k_1 is the proportionality factor connecting the specific and the absolute values of the core voltage.

The specific operating current of the core, $aw_\sim$, is computed from the known absolute current $I_\sim$:

$$aw_\sim = \frac{w_\sim C_2}{l_{av}} I_\sim = k_2 I_\sim, \tag{2}$$

where l_{av} is the average length of the magnetic circuit's lines of force, in cm, C_2 is a coefficient which indicates what portion of the total effective phase current of the core comprises the effective current of one working winding (for the circuit of Fig. 1, $C_2 \approx 0.71$), k_2 is the proportionality factor between the specific and absolute values of core operating current.

The specific volt-ampere characteristics are constructed for various fixed values of the resulting specific control current aw_{cr}:

$$aw_c = \sum_{N=1}^{N=n} \frac{w_{cN}}{l_{av}} I_{cN} = \sum_{N=1}^{N=n} k_{3N} I_{cN} = \sum_{N=1}^{N=n} aw_{cN}, \tag{3}$$

where w_{cN} is the number of turns of the N'th control winding, I_{cN} is the control current in the N'th control winding and K_{3N} is a proportionality factor between the specific and absolute values of control current.

The coefficients k_1, k_2 and k_{3N} are determined by the parameters of the saturable core and of the circuit for the loaded windings and, for a given saturable core and circuit, are constants.

The volt-ampere characteristics of the saturable core load (diode bridge and motor) must be given in specific units in accordance with equations (2) and (3), and are constructed in the fourth quadrant (Fig. 2). Here, the axis of the specific operating current is common to the volt-ampere characteristics of the saturable core and of the load.

The specific volt-ampere characteristics of the load are conveniently constructed by individual component. The specific volt-ampere characteristics of the froward impedance of the diode bridge may be constructed,

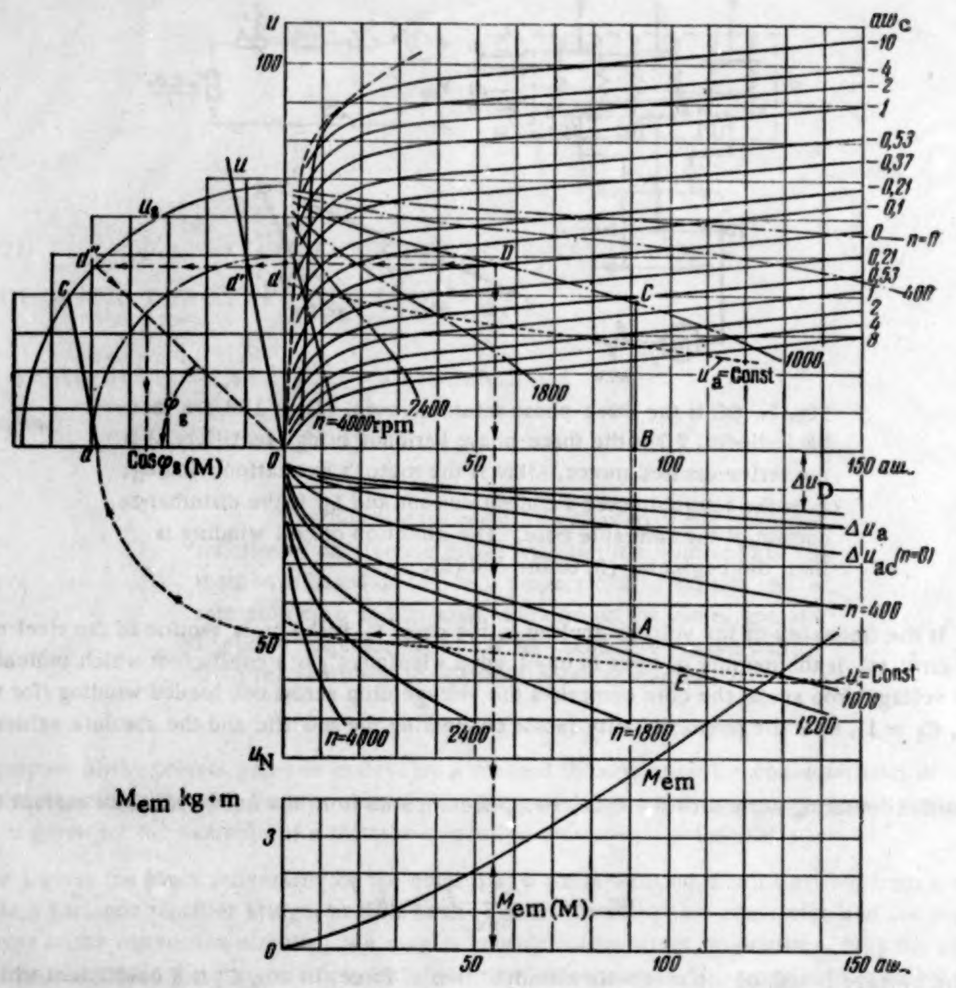


Fig. 2. Specific saturable core (SC) volt-ampere characteristics, together with the trajectories of the operating point (first quadrant) and the specific volt-ampere characteristics of the load (fourth quadrant). The dependence of M_{em} on the specific current, $aw_\sim$ and the graphic method for determining the characteristics of a throttle sysetm.

for example, from the volt-ampere characteristics of the diode bridge. If the forward voltage drop across the interior impedances of the diode bridge ΔU_D leads to the phase value, but the current $I_\sim$ is measured in the bridge's linear conductor, then the specific internal voltage drop across the bridge's forward impedance is determined from the formula

$$\Delta u_D = k_1 \Delta U_D. \tag{4}$$

The specific current of the bridge is determined from the expression

$$aw_\sim = k_2 I_\sim. \tag{5}$$

In formulae (4) and (5), ΔU_D is the internal forward voltage drop in the bridge and $I_\sim$ is the current in the bridge's linear conductor.

The dc motor is connected on the side of the rectified current. To construct its specific volt-ampere characteristics, it is necessary that the rectified voltage and rectified current be brought, respectively, to the phase value of the voltage supplied to the diode bridge (with attenuation factor γ_1), and to the current in the

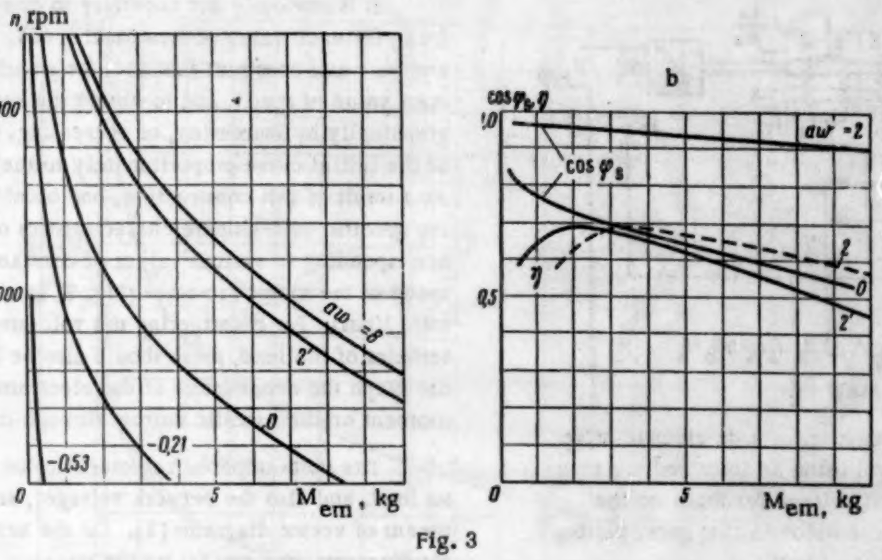


Fig. 3

supplying conductor (with attenuation factor γ_2). For a three-phase Larionov bridge, one can take $\gamma_1 = 0.427$ and $\gamma_2 = 0.82$ [3]. In a first approximation, the attenuation factors are considered independent of the states of the saturable reactor and the load. It is also convenient to construct the volt-ampere characteristics of the motor by its components.

The abscissa points of the specific volt-ampere characteristics of the motor's armature circuit are determined from the equation

$$aw_{\sim} = k_2\gamma_2 I_a, \tag{6}$$

where I_a is the current (rectified) in the armature circuit.

The ordinate of the specific volt-ampere characteristics of the internal impedance of the armature circuit is determined by the expression

$$\Delta u_{ac} = k_1\gamma_1 r_{ac} I_a. \tag{7}$$

To construct the specific volt-ampere characteristics of the armature circuit impedance, it is sufficient to determine the abscissa and ordinate of one point, to draw a supplementary line through this point and the origin of coordinates, and to add, graphically, the ordinates of the line to the ordinates of the curve for Δu_D.

There is an analogous construction for the component of the voltage drop in the brush contacts which, in many practical cases, is taken to be constant.

To construct the specific volt-ampere characteristics of the rotational counter emf, u_r, taking the armature reaction into account, it suffices to have the characteristics of the free-running series motor, with armature reaction taken into account. For definiteness, we assume that the curve for free-running is given in the form of the dependence of the product of the flux of one pole by a design constant of the machine, C_e, on the armature current.

The curve for the specific rotational emf is constructed, for constant speed, by points based on the free-running characteristics and on the equation

$$u_r = k_1\gamma_1 C_e \Phi n, \tag{8}$$

where n is the speed of motor shaft rotation, in rpm.

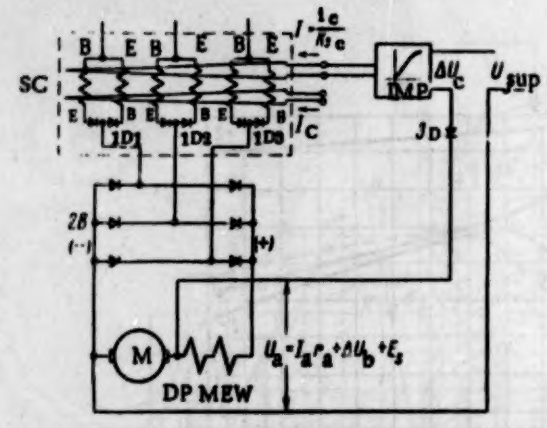


Fig. 4. Block schematic for a dc electric drive with throttle control using an intermediate magnetic amplifier and voltage feedback on the armature. 3D is a semiconducting gate, passing current only when $U_{sup} > U_a$.

It is obviously not necessary to compute the curves for u_r for each value of rotational speed. It suffices to compute and construct just one, for an arbitrary constant value of speed, and to obtain the remaining ones graphically by increasing, or decreasing, the ordinates of the initial curve proportionately to the ratio of speeds. As a result of this construction, one obtains the resulting specific volt-ampere characteristics of the SC load, corresponding to various values of constant rational speed of the motor armature (Fig. 2, fourth quadrant, solid lines). For constructing the volt-ampere characteristics of the load, there should also be built up on the graph the dependence of the electromagnetic moment on the specific current through the SC load.

The volt-ampere characteristics of an SC and its load, and also the network voltages, are related by means of vector diagrams [1]. On the basis of the vector diagrams, one can set up the equation for the equivalent sinusoids of the voltage:

$$U_{np}^2 = (U_L + U_{sat}\cos\varphi_{sat})^2 + U_{sat}^2\sin^2\varphi_{sat}$$

After some transformations, this equation can be written in the form

$$\frac{U_{sat}^2}{U_{np}^2} + \frac{I_\sim^2}{\left(\frac{U_{np}}{R_e}\right)^2} + 2\,\frac{U_{sat} I\cos\varphi_{sat}}{\left(\frac{U_{np}}{R}\right)U_{np}} = 1, \qquad (9)$$

where $R_e = r_D + \gamma_1 r_{ac}$ $\gamma_2 + \gamma_1 C_e \Phi_n$ γ_2 l_2, U_{sat} and U_{np} are, respectively, the voltages on the core and in the network, I_a is the current in the motor's armature circuit and $I_\sim$ is the core's working current (or the current supplied to the motor's armature circuit).

If the numerator and denominator in equation (9) are multiplied by, respectively, k_1 and k_2, one obtains an equation in the same specific units as the volt-ampere characteristics of the SC.

If, in equation (9), we set R_e equal to a constant, and ignore the active losses in the SC ($\cos\varphi_{SC} = 0$), we shall have obtained the generally known equation for the trajectory of the SC's operating point (abbreviated below to OPT) in the form of the canonical equation for an ellipse, by which the OPT is generally found by a computational method. When an SC operates in a motor, the resistance R_e is changing and nonlinear, and to ignore the active losses in the SC leads to a significant error in the determination of the coefficients of power and electric-drive efficiency. The facts just mentioned predetermine the great difficulty which accompanies attempts to use the most widely disseminated methods for computing characteristics by means of OPT's in the form of ellipses. A simpler, and more accurate, solution may be found with the use of graphic constructions of OPT's and electric drive characteristics.

In the construction of OPT's, a circle is used whose radius is $m_u u_c$, which is the geometric locus of the ends of the vectors of the specific network voltage, with the assumption being made that the network voltage is constant (Fig. 2, second quadrant). Here, m_u is the voltage scale on the u axis. We assume that the current vector coincides with the horizontal radius of the circle u_c.

Let it be required to find the ordinate of the OPT point which corresponds to point A on the specific volt-ampere characteristic for load with n = 1000 rpm (Fig. 2). It is necessary for this that the segment AB be translated to the horizontal radius of circle u_c (Fig. 2, second quadrant) and laid off from the center 0 (point a).

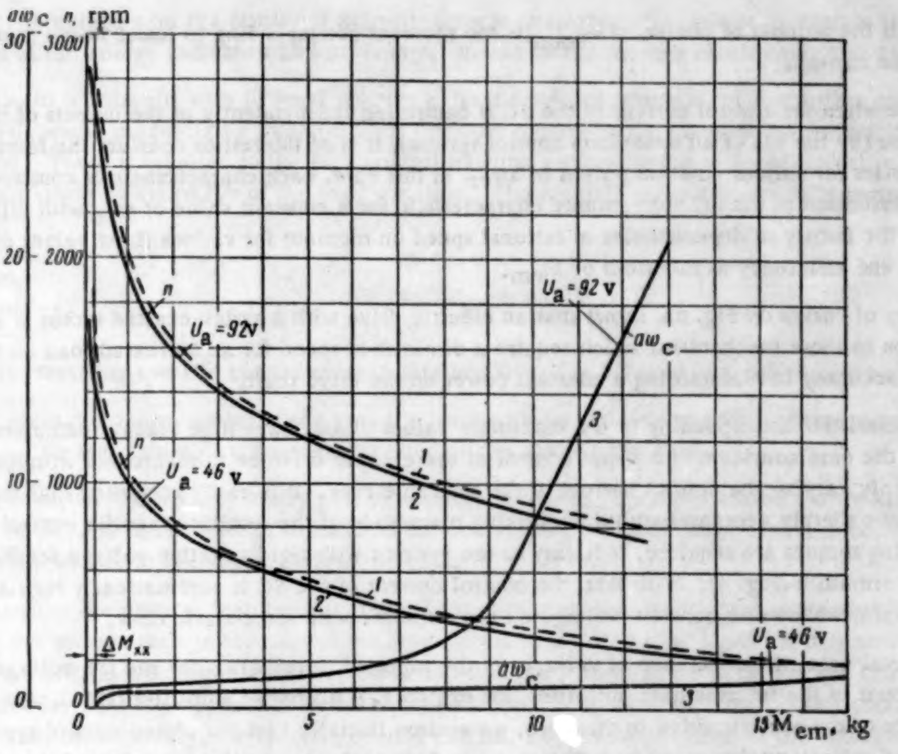


Fig. 5. Mechanical characteristics of the motor. 1) is without taking into account the voltage drop in the circuit of the amplifier control winding, 2) does takes it into account; 3) is the dependence of the specific control current on the torque, M_{em}, with specific displacement current $aw_{di} = -1$.

Assuming, for definiteness, that the losses in the SC are characterized by a constant angle φ_{SC} [4], we draw the line ac at the angle φ_{SC} to the horizontal radius. The segment ac = BC is the ordinate of the desired point of the OPT. The abscissa of the OPT point is the point B on the $aw_{\sim}$ axis. The ordinate of the OPT is sought, essentially, on the vector diagram [1], where one can readily convince oneself whether or not the point c is joined to the origin of coordinates (Cf. also Fig. 1).

As the point A moves along the volt-ampere characteristic of the load, the segment BC also moves, and changes its magnitude. With this, the point C traces out the OPT corresponding to the given volt-ampere characteristic of the load. The family of volt-ampere characteristics of the load corresponds to the family of OPT's (Fig. 2, first quadrant, dash lines). It is characteristic that the OPT's thus obtained are not ellipses. This arises from the taking into account of the nonlinearity of the load and of the active component of the voltage drop across the SC.

The point of intersection of an OPT with an SC volt-ampere characteristic, for example, point D, determines itself the specific resulting value of the SC's control current, the speed of rotation of the motor shaft, the specific current of the working winding of the SC, the motor's electromagnetic moment, M_{em}, and the coefficient of system power, $\cos\varphi_s$. $\cos\varphi_s$ is determined graphically. For this, point D is projected on the u axis (point d) and translated to the line u (Od = Od'). From point d' a horizontal line is drawn up to the intersection with circle u_c (point d"). The angle between the line Od" and the horizontal radius corresponds to the angular shift between the current and voltage of the circuit. Computing $\cos\varphi_s$ is conveniently done by using an additional circle of unit radius.

The efficiency is computed on the basis of the quantities obtained above by means of the formula

$$\eta_s = \frac{P_2}{P_s} = \frac{(M_{em} - \Delta M_{cons})\, n k_2}{0.975 m_1 U_{np}\, aw_{\sim} \cos\varphi_s}. \tag{10}$$

Here, m_1 is the number of phases, ΔM_{const} is the moment corresponding to losses in the motor which do not depend on the current.

In the case when the control current of the SC is controlled independently of the outputs of the motor or the saturable core (by the use of an open-loop control system), it is of interest to consider the family of electric drive characteristics for various constant values of aw_{cr}. In this case, each characteristic is constructed by using the points of intersection of the SC volt-ampere characteristic for a constant value of aw_c with different OPT's. Figure 3a shows the family of dependencies of rational speed on moment for various fixed values of aw_{cr}. Fig. 3b, shows $\cos\varphi_s$ and efficiency as functions of M_{em}.

The family of curves on Fig. 3a, shows that an electric drive with a series-excited motor is advantageously used for operation in those mechanisms which require a diminished speed for an increased load on the shaft, or a low degree of accuracy in maintaining a constant power on the drive shaft.

The characteristics corresponding to the maximum values of aw_c have high energy indicators (Fig. 3b). It is essential in the case considered here that control of the electric drive be implemented without automatic regulators by simply varying the control current of the saturable core. It is easily remarked that the magnitude of the initial torque sharply decreases with a decreasing magnitude of the resulting specific control current aw_{cr}. When large starting torques are required, it is easy to use systems with rigid negative voltage feedback, for example, on the armature (Fig. 4). With this, the control current of the SC is automatically regulated, and guarantees the maintenance of a constant voltage on the armature with some static error.

In the general case, the difference of voltages of the supplied (forcing) signal and the voltage feedback is supplied to the input of the intermediate amplifier, for example, a magnetic amplifier (IMA). For computing the characteristics of the electric drive in this case, we assume initially that the closed control system provides ideal constancy of voltage on the armature. This is equivalent to the assumption that the voltage drop in the circuit of the control winding of the IMA, ΔU_C, equals zero. In the steady state, the voltage on the armature is comprised of the voltage drops in the armature impedance, the brush contacts and the rotational counter emf E:

$$U_a = I_a r_a + \Delta U_b + E = \text{const.} \tag{11}$$

The rectified voltage at the output of rectifier 2D (Fig. 4) must increase with increasing current in order to compensate for the voltage drops across the impedances of the secondary and principal pole windings. The resulting volt-ampere characteristics for the SC load, corresponding to a constancy of voltage on the motor armature, are easily constructed if lines are drawn equally spaced from the volt-ampere characteristic of the sum of the diode bridge impedance and the secondary and principal pole windings, i.e., at equal distances from the sum of the volt-ampere characteristics of those components of the load impedance which are shunted by the voltage feedback path (Fig. 2, fourth quadrant, dotted line).

The mechanical characteristics of an electric drive can be constructed by using the points of intersection of the lines of constant voltage on the armature, u_a = const, with the motor's volt-ampere characteristics corresponding to a constant speed of rotation of the motor's armature, and the curve of M_{em} as a function of $aw_\sim$. Indeed, the point E (Fig. 2) itself determines the motor's speed and torque. The set of points of intersection of the lines u_a = const with the motor's volt-ampere characteristics gives the motor's mechanical characteristics (Fig. 5, 1).

The method presented for constructing the characteristics of an electric motor from its volt-ampere characteristics has independent value, and may be recommended for calculating the characteristics of motors with excitation fluxes which depend nonlinearly on armature current.

The further calculations have the purpose of obtaining initial data for selecting the intermediate amplifier and for designing the control system and a correcting device for the mechanical characteristics, account being taken of the impedance in the circuit of the IMA's control winding. For this, it is necessary to construct the OPT's by load volt-ampere characteristics which correspond to the constancy of the voltage on the armature. The method of constructing the OPT's does not differ essentially from the case considered above; it necessarily follows that point A slips along the line u_a = const (Fig. 2).

The OPT, combined with the family of SC volt-ampere characteristics, makes it possible to construct the dependence of the energy indicators and SC control current on the motor's electromagnetic torque.

Ordinarily, to SC circuits with internal positive current feedback, there is fed a negative constant displacement signal, whose magnetizing force (m.f.) is directed counter to the constant component of the m.f. of the load winding. There is supplied to the SC's control winding a signal whose m.f. coincides, in action, with the constant component of the m.f.'s of the load windings. The specific current in the SC's control winding, aw_c, will equal

$$aw_c = aw_{cr} + aw_{dl}, \tag{12}$$

where aw_{cr} is the resulting specific control current obtained from the SC's specific volt-ampere characteristics.

Figure 5 gives the dependencies of the speed, $\underline{n}$, (dotted lines) and the specific control current, aw_c, on the torque, M_{em}, for two values of voltage maintained on the motor's armature.

By using equation (3), we may easily pass to absolute values of SC control current, if we know the design parameters of the SC and the number of turns in its control winding. A knowledge of the absolute values of SC control current allows one to choose, or design, the IMA.

Let us consider the general features of a method for correcting the mechanical characteristics with account being taken of the voltage drop in the control winding circuit of the IMA (CW IMA). If it is assumed that the impedance of the control winding circuit is constant and that the current at the output of the IMA (control current of the saturable core) is proportional to the control current of the IMA, then the control current of the saturable core and, consequently, the specific control current, aw_c, is proportional to the voltage drop in the control winding circuit of the IMA.

If the armature circuit is invariable, the voltage drop across the IMA's control winding can only be obtained as the result of decreased emf of armature rotation:

$$\Delta U_c = E - E_s. \tag{13}$$

Here, and below, the subscript $\underline{s}$ denote those quantities obtained when account is taken of the voltage drop, ΔU_c, in the IMA's control winding circuit.

From equation (13), an expression is easily obtained for the deviation of the speed engendered by the voltage drop in the circuit of the CW IMA. If the magnitude of the voltage drop in the CW IMA is known for one of the values of specific control current of the saturable core, aw_{kn}, then the deviation of rotational speed can be found from the formula

$$\Delta n_y = n - n_s = \frac{aw_c}{aw_{kn}} \frac{\Delta U_{ky6} n}{E}.$$

The deviation of speed, Δn_c, can be found by using the specific volt-ampere characteristics:

$$\Delta n_c = \frac{aw_c}{aw_{kn}} \frac{\Delta U_{kn} k_1 \gamma_1 m_u n}{m_u u_r}, \tag{14}$$

where m_u is the voltage scale on the $\underline{u}$ axis of the specific volt-ampere characteristics for load (throttle), $m_u u_r$ is the specific value of the rotational counter emf in the scale of the volt-ampere characteristics (Fig. 2), in mm.

The speed of rotation, $\underline{n}$, is determined from the graphs. By formula (14), all values of the current quantities are substituted in the graphs of Figs. 2 and 5 for constant values of M_{em}. The obtained value of Δn_c must be computed from the curves $n = f(M_{em})$ for those same values of M_{em}. A corrected curve for $n = f(M_{em})$ is shown in Fig. 5 (curve 2).

The circuit of Fig. 4 is recommended for use in those cases when the requirement for increasing starting torque is decisive.

All graphs and characteristics provided in this paper refer to an actual type PN-68 motor, in which the independent excitation winding is replaced by a series one. The nominal values for such a motor are: 110 volts, 42 amperes, speed of rotation 1850 rpm, torque on shaft 2.4 kilograms, armature impedance when heated, 0.146 ohms.

The magnetic circuit of the three-phase saturable core is made of pi-shaped disks with the average length of the lines of force l_a = 45 cm, and has an active cross section S_c = 17.7 cm^2. The type of steel is É-310, with heat-treating after stamping. The number of turns in one load winding is $w_{\sim}$ = 120.

The type ABC selenium diodes, with bead dimensions of 100 × 100, are cooled by blowers. With forced cooling (the basic variant) seven beads are connected in series and three in parallel in the arms, and with natural cooling, seven in series and seven in parallel. The line voltage is 127 volts.

In Fig. 3, the characteristics drawn in solid lines refer to the variant with forced cooling of the selenium diodes, the dotted curves correspond to the natural cooling variant.

SUMMARY

1. In terms of a series-excited electric motor, a new graphical method was presented for constructing the characteristics of dc electric drives with throttle control in the armature circuit, with account taken of nonlinearities, in the cases when:

a) the control current for the saturable core is controlled independently of the system's output quantity;

b) the control current is controlled in dependence on voltage, for example, that on the motor armature.

2. As a result of the analysis of the characteristics of series-excited dc electric motors with throttle control, it was established that application of the electric-drive system cited is particularly advantageous in those cases which require a decrease in speed when the load on the motor shaft increases, or which require the maintenance of constant power on the shaft with a low degree of accuracy.

The method presented may be useful for designers and investigators working in the domain of the electric drives referred to.

LITERATURE CITED

[1] V. S. Kulebakin, "On the use of semiconducting diodes in automated electric drive systems," [in Russian]. Izv. AN SSSR, Otd. Tekhn. N. 2 (1958).

[2] M. G. Chilikin, M. M. Sokolov and V. I. Klyuchev, "Controlled asychronous electric drives with saturated cores and excitation motors," [in Russian]. Élektrichestvo 1 (1956).

[3] I. L. Kaganov, Electronic and Ionic Transformers [in Russian]. Gosénergoizdat (1950).

[4] D. Alenchikov, "Computing the static characteristics of throttle-controlled systems," [in Russian]. Automation and Remote Control (USSR) 20, 5 (1959).

Received March 12, 1959

ON DETERMINING THE FEEDBACK PARAMETERS FOR A VIBRATION REGULATOR OF ELECTRIC MOTOR SPEED

L. L. Rotkop
(Odessa)

A method is presented for determining the parameters of a stabilizing feedback loop of a vibration regulator of electric motor speed. The peculiarities of a contactless relay feedback loop are considered. Examples are given of the determination of the parameters of the stabilizing feedback loop of a vibration regulator.

The normal mode of operation of a vibration regulator of electric motor rotational speed (VR) is the autooscillatory mode. One of the indicators of the quality of a control system with VR is the magnitude of the autooscillation of speed about its mean value. In designing a VR, one ordinarily tends to increase the frequency and decrease the amplitude of the speed autooscillations to the technically possible limits. This sometimes succeeds with changing the parameters of the linear portion and the relay element, but ordinarily the possibilities of increasing autooscillation frequency at the costof changing the system parameters are limited. For stabilizing relay systems, broad use is made of internal feedback paths which shunt the relay element [1, 2]. This paper is devoted to a method for determining the parameters of such feedback paths.

The characteristic of a relay system

$$J(\omega) = -\frac{1}{\omega}\dot{\tilde{z}}\left(\frac{\pi}{\omega}\right) - j\tilde{z}\left(\frac{\pi}{\omega}\right), \tag{1}$$

the determination of which is given in [1], is very convenient for the qualitative and quantitative investigation of the effect of the parameters of an internal feedback path on the frequency of autooscillation of a relay system with symmetric relay characteristics with no dead zones.

In formula (1), $\tilde{z}(\pi/\omega)$ and $\dot{\tilde{z}}(\pi/\omega)$ are the output quantitity of the linear portion and its derivative (where $t = \pi/\omega$) when the input is pulsed by symmetric, sign-alternating pulses of constant height. The tilde over the symbols indicates the periodic nature of their variation.

In fact, the autooscillation frequency of a relay system is determined by the intersection of the characteristic $\mathrm{Im}J(\omega)$ with the line $\mathrm{Im}J(\omega) = -\kappa_0$ for $\mathrm{Re}J(\omega) < 0$ [1], and the relay system characteristic may be presented as the sum of the characteristics of a relay system without feedback, $\mathrm{Im}J_1(\omega)$, $\mathrm{Re}J_1(\omega)$, and a relay system in which the linear portion is a feedback path, $\mathrm{Im}J_2(\omega)$, $\mathrm{Re}J_2(\omega)$.

The construction of the aforementioned characteristic gives a graphic presentation of the servicability of the feedback structure generally, or of the values of the parameters of the feedback path for obtaining the necessary frequency of regulator autooscillation. The determination of the autooscillation frequency of relay systems with asymmetric characteristics, among which are VR systems, reduces to the determination of two parameters: the frequency ω_p and the relative switching time $\gamma = t_1/(t_1 = t_2)$, where t_1 is the pulse duration and t_2 is the pause duration.

In this case the solution of the problem is more complex since the geometric construction is less revealing than in the case of symmetric characteristics, and does not always allow the proper conclusions as to choice of feedback parameters to be made.

Application of Analytical Methods for Symmetric Relay Systems to Vibration Regulators of Electric Motors

The block schematic of a VR circuit and the characteristic of its relay element are shown in Fig. 1a. In Fig. 2, as examples are shown VR circuits using contact control apparatus (Fig. 2a) and contactless control apparatus – a magnetic amplifier and a magnetic contactless relay (Fig. 2b) [3]. The frequency of speed autooscillation, ω_p, and the relative switching time, γ, in a system with VR depend, under almost equal conditions, on the size of the load on the motor [1, 4]. As the motor load varies, the autooscillation frequency varies parabolically within quite wide limits, and γ varies from 0 to 1. The choice of feedback parameters for increasing autooscillation frequency can be made for a given motor load (γ = const), and then the effect of the chosen feedback parameters on autooscillation frequency can be investigated over the entire range of load variation. If the value of the load is not chosen arbitrarily, but is so chosen that the equality $\gamma = 0.5$ ($t_\gamma = t_2$) holds, then the speed autooscillations will be symmetric about the axis $z = z_{1av}$, where

$$z_{1\,av} = \frac{1}{2}\left[\tilde{z}_{1N}\left(\frac{T}{2}\right) + \tilde{z}_{1N}(T)\right], \tag{2}$$

where $\tilde{z}_{1N}$ is the output quantity of the linear portion (Fig. 1a) and $T = t_1 + t_2$ is the period of autooscillation.

By the proper choice of coordinate axes for the autooscillations of the system being considered, for $\gamma = 0.5$, we can be brought to a relay system with symmetric relay characteristics. Figure 1b, gives the block schematic of such a system after translation of the coordinate axes. For this system, the following relationships are valid

$$\tilde{y} = \tilde{y}_1 - f_s(n), \tag{3}$$

$$\tilde{z}_1 = \tilde{z}_{1_N} - z_{1_{av}}, \tag{4}$$

$$\tilde{x} = -\tilde{z}_1 - \tilde{z}_2 = -[\tilde{z}_2 + (\tilde{z}_{1_N} - z_{1_{av}})], \quad \tilde{z}_1(t) = -\tilde{z}_1\left(t + \frac{T}{2}\right), \tag{5}$$

$$\tilde{z}_2(t) = -\tilde{z}_2\left(t + \frac{T}{2}\right).$$

Here, $f_s(n)$ is the magnitude of the VR load for $\gamma = 0.5$; for the remaining nomenclature, Cf. Fig. 1.

The system's block schematic for $\gamma = 0.5$ (Fig. 1b) corresponds to a relay system for automatic control with symmetric relay characteristics. In this case one may use the simplest method of determining the feedback parameters by means of the relay system's characteristic $J(\omega)$, which was cited above.

For $\gamma \pm 0.5$, the autooscillations in the system have a nonsymmetric character ($t_1 \pm t_2$). The condition for relay transfer (Fig. 1) have the form:

$$\tilde{z}_1\left(\gamma\frac{2\pi}{\omega}\right) + \tilde{z}_2\left(\gamma\frac{2\pi}{\omega}\right) = x_0, \qquad \tilde{z}_1\left(\frac{2\pi}{\omega}\right) + \tilde{z}_2\left(\frac{2\pi}{\omega}\right) = -x_0. \tag{6}$$

The linear portion of the system (the electric motor) is a low-frequency filter, and therefore the increment of the output magnitude of the linear portion decreases as the frequency of the controlling stimulus $\underline{y}$ (Fig. 1b) increases. Consequently, when there is a correcting signal in the feedback path, the switching conditions in (6) hold for higher frequencies of the controlling stimulus $\underline{y}$ only when the increment of the linear portion's output signal coincides in sign with the increment of the feedback signal. The speed increment of the motor (in the motor mode of operation) is always positive at the end of a pulse (acceleration) and negative at the end of a pause (braking). The conditions under which the feedback increases the autooscillation frequency are written in the form

$$\tilde{z}_2\left(\gamma\,\frac{2\pi}{\omega}\right) > 0,\quad \tilde{z}_2\left(\frac{2\pi}{\omega}\right) < 0. \tag{7}$$

To determine the sign of the feedback path's output quantity, it is necessary to investigate the value of this quantity when the input receives an asymmetric sequence of pulses. In this case, the controlling stimulus y can be presented in the form of a constant stimulus and a succession of asymmetric sign-alternating pulses of constant height $k_p = (y_A - y_B)/2$. The reactions of the feedback path at the ends of pulses and pauses to an asymmetric succession of pulses may be determined from the expressions [1]

$$\tilde{z}_{2N}\left(\gamma\,\frac{2\pi}{\omega}\right) = z_{av} + k_p \sum_{s=1}^{n} \frac{P(p_s)}{Q'(p_s)\,p_s}\,\frac{e^{p_s(0.5-\gamma)\gamma\frac{2\pi}{\omega}} - e^{p_s\gamma\frac{2\pi}{\omega}}}{1 + e^{p_s\gamma\frac{2\pi}{\omega}}}, \tag{8a}$$

$$\tilde{z}_{2N}\left(\frac{2\pi}{\omega}\right) = z_{av} + k_p \sum_{s=1}^{n} \frac{P(p_s)}{Q'(p_s)\,p_s}\,\frac{e^{p_s(0.5-\gamma)\frac{2\pi}{\omega}} - e^{p_s\frac{2\pi}{\omega}}}{1 + e^{p_s\frac{2\pi}{\omega}}}, \tag{8b}$$

where $P(p)/Q(p)$ is the transfer function of the feedback path.

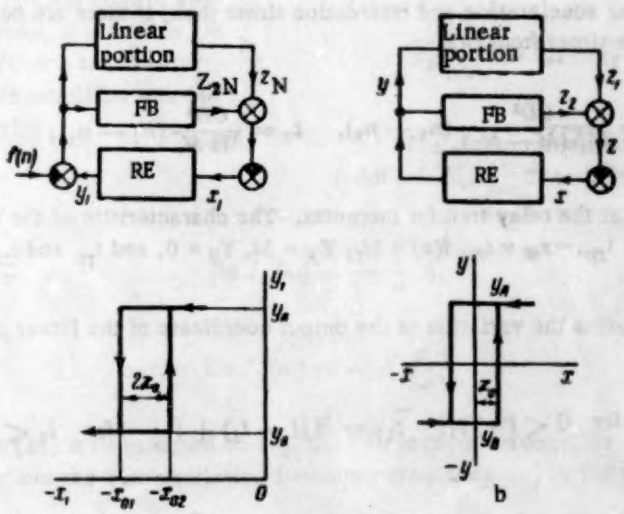


Fig. 1

We shall consider only the second term in each of equations (8), the ones expressing the increment of the feedback path's output quantity. For $\gamma = 0.5$, these terms differ from each other only in sign and, as the frequency varies from 0 to ∞, equal the imaginary part of the characteristic $J_2(\omega)$ [Cf. (1)]. The signs of the second terms in (8), for negative real nonzero poles of the feedback path's transfer function, do not depend on the magnitude of γ, since always, as γ varies from 0 to 1, in the expressions (8),

$$\exp p_s(0{,}5 - \gamma)\frac{2\pi}{\omega} > \exp p_s\frac{2\pi}{\omega}.$$

Consequently, if the transfer function of the feedback path has simple real negative poles, conditions (7) hold for all values of γ, and the parameters of the VR's feedback path, selected for $\gamma = 0.5$, increase the auto-oscillation frequency for all values of the load ($0 \le \gamma \le 1$). In those cases when there are complex poles, and for $\gamma = 0.5$ and for certain values of controlling stimulus frequency the second terms in (8) change sign (the

characteristic $\mathrm{Im}J_2(\omega)$ crosses the axis of abscissas), it is necessary to determine, from expressions (8), the values γ_1 and γ_2 for which the second terms in (8) equal zero. In this case, conditions (7) hold in the range of load variation $\gamma_1 \le \gamma \le \gamma_2$, and the feedback chosen for $\gamma = 0.5$ increases the autooscillation frequency only for changes of motor load in that range. We shall not consider more complex feedback paths in this paper.

Thus, by using the method for determining feedback parameters for symmetric relay systems, one can determine the feedback parameters for a VR with $\gamma = 0.5$ and then, by the use of conditions (7), one can learn in which range of load variation this feedback will increase the autooscillation frequency.

As an example, let us find the feedback parameters for a relay system of asynchronous motor speed control (Fig. 2a).

We write the motor equations in the form

$$M - M_r = \frac{GD^2}{375}\frac{dn}{dt} \quad \text{for} \quad 0 < t < t_1, \quad -M_r = \frac{GD^2}{375}\frac{dn}{dt} \quad \text{for} \quad t_1 < t < T, \tag{9}$$

where M and M_r are, respectively, the driving torque and the resistance torque on the motor shaft, GD^2 is the moment of gyration of the drive and n is the motor speed.

We take the tachometer generator equation in the form $i = k_1 n$, where i is the current in the control relay circuit.

If we assume that the amplitude of the speed autooscillation is sufficiently small so that we may consider M = const and M_r = const during acceleration and retardation times then, if there are no lags in the relay, we may obtain the pulse and pause times from (9):

$$t_1 = \frac{GD^2}{375\,(M - M_r)}(n_1 - n_2), \quad t_2 = \frac{GD^2}{375\,M_r}(n_1 - n_2), \tag{10}$$

where n_1 and n_2 are the speeds at the relay transfer moments. The characteristic of the relay element is given in Fig. 1a, where $x_1 = i$, $-x_{a1} = i_{rp}$, $-x_{a2} = i_{rr}$, $f(n) = M_r$, $Y_A = M$, $Y_B = 0$, and i_{rp} and i_{rr} are the pickup and reset currents of the relay.

The equation which describes the variation of the output coordinate of the linear portion is found from (9) and (10) in the form

$$\tilde{z}_N = At + i_{rr} \quad \text{for} \quad 0 < t < t_1, \quad \tilde{z}_{1N} = B(t - t_1) + i_{rp} \quad \text{for} \quad t_1 < t < T, \tag{11}$$

where

$$A = \frac{375\,(M - M_r)}{GD^2}k_1 \text{ and } \qquad B = \frac{-375\,M_r}{GD^2}k_1.$$

The relative time for switching on is found from (11)

$$\gamma = \frac{M_r}{M}. \tag{12}$$

The frequency of speed autooscillation is defined by the formula

$$\omega = \frac{2\pi\, 375\, M}{GD^2\,(n_1 - n_2)}(\gamma - \gamma^2). \tag{13}$$

The symmetric autooscillations about the axis $z_{1av} = (i_{rp} + i_{rr})/2$ occur for $f_s(n) = M/2$ or $M_r = M/2$ ($\gamma = 0.5$).

We now turn to the symmetric relay system (Fig. 1b) for which, on the basis of (3)-(5), we may write the relationships

$$f_s(n) = \frac{1}{2} M,$$

$$y_A = \frac{1}{2} M,$$

$$y_B = -\frac{1}{2} M,$$

$$\tilde{z}_1 = At - \varkappa_0$$

$$\left[\varkappa_0 = \frac{1}{2}(i_{rp} - i_{rr})\right],$$

bearing in mind that

$$\tilde{z}_1\left(\frac{\pi}{\omega}\right) = \varkappa_0, \quad \tilde{z}_1 = A\left(t - \frac{\pi}{2\omega}\right). \tag{14}$$

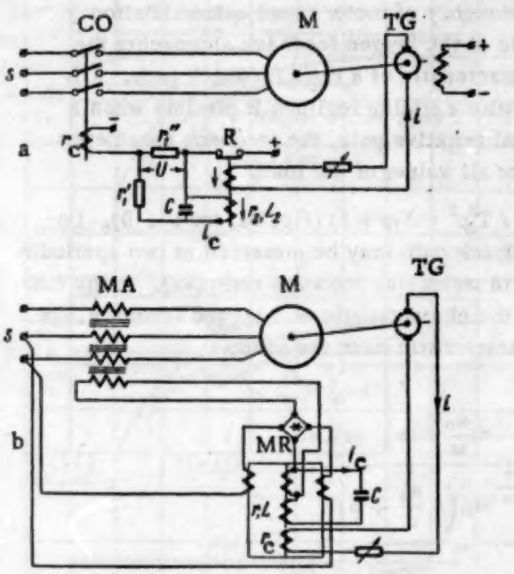


Fig. 2. a) is a vibration speed regulator for an asynchronous motor: $r_1 >> r_c$, $r_2 >> r_c$, $r_1' + r_1''$; b) is the same but with contactless control apparatus used: CO is a contactor, R is a relay, M is the asychronous motor, TG is a tachometer generator, MA is a magnetic amplifier and MR is a magnetic contactless relay.

The characteristic of the relay system without feedback ($\tilde{z}_2(t) = 0$) is defined by the expressions

$$\operatorname{Re} J_1(\omega) = -\frac{1}{\omega} A, \tag{15}$$

$$\operatorname{Im} J_1(\omega) = -A \frac{\pi}{2\omega}. \tag{16}$$

The characteristic in (16) is constructed on Fig. 3. The point of intersection of this characteristic with the line $\operatorname{Im} J_2(\omega) = -\kappa_0$ defines the autooscillation frequency frequency ω_{r1} in the system without feedback for $\gamma = 0.5$.

We now construct the characteristic of a relay system in which the linear portion is a feedback path ($\tilde{z}_{1N}(t) = 0$, $z_1 = z_{1av}$).

1. We consider two cases. The feedback element has the transfer function $W(p) = k/(1 + T_1p)$ (delayed feedback). A circuit for a system with a vibration regulator of electric motor speed with lagged feedback introduced is shown in Fig. 2a (we assume that $L_2 = 0$). After a translation of coordinates (Fig. 1b), the input of the feedback path is acted upon by sign-alternating voltages pulses of height $k_p = U/2$ (here and in the sequel, the input to the feedback circuit will be denoted by $\tilde{y} = U/2$, and its output quantity is taken to be the current in the winding of the feedback stabilization relay, $\tilde{z}_2 = i_k$).

By bearing in mind that $i_k(0) = -\kappa_0$ and $i_k(\pi/\omega) = \kappa_0$, we obtain

$$\operatorname{Re} J_2(\omega) = -\frac{U}{2(r_1 + r_2) T_1 \omega}\left(1 - \operatorname{th} \frac{\pi}{2\omega T_1}\right), \tag{17}$$

$$\operatorname{Im} J_2(\omega) = -\frac{U}{2(r_1 + r_2)} \operatorname{th} \frac{\pi}{2\omega T_1}. \tag{18}$$

The characteristic in (18) is given in Fig. 3a. The characteristic of the relay system, $\mathrm{Im}J(\omega) = \mathrm{Im}J_1(\omega) + \mathrm{Im}J_2(\omega)$ (Fig. 3a) shows that the lagged feedback increases the frequency of motor speed autooscillation ($\omega_{p_2} > \omega_{p_1}$). As the time constant, T_1, decreases, the characteristic of the lagged feedback approaches the straight line $\mathrm{Im}J_2(\omega) = -U/2(r_1 + r_2)$, which corresponds to the characteristic of a rigid feedback path. It is necessary that the inequality $U/2\,(r_1 + r_2) < \kappa_0$ be observed, otherwise a sliding regimen is possible when a rigid feedback path is used. Since the transfer function has one real negative pole, the feedback parameters chosen for $\gamma = 0.5$, increase the speed autooscillation frequency for all values of the load.

2. The feedback element has the transfer function $W(p) = k / T_0^2p^2 + T_1p + 1)$ (Fig. 2a, for $L \pm 0$). Depending on the form of the characteristic equation's roots, the feedback path may be presented as two aperiodic links in series (roots are real and equal) or as two oscillatory links in series (the roots are complex). In the first case, the characteristic of such a feedback path differs little from the characteristic of a lagged feedback. In the second case ($p_{1,2} = -b \pm j\omega_0$), the expressions for the relay characteristic have the form:

$$\mathrm{Re}\, J_2(\omega) = -\frac{Ue^{-b\frac{\pi}{\omega}}\sin\alpha}{\omega\omega_0(r_1+r_2)}\,\frac{\sin\left(\alpha - \pi\frac{\omega_0}{\omega} - \varphi\right)}{\sin\varphi + e^{-b\frac{\pi}{\omega}}\sin\left(\pi\frac{\omega_0}{\omega} + \varphi\right)}, \tag{19}$$

$$\mathrm{Im}\, J_2(\omega) = -\frac{U}{2\,(r_1+r_2)}\left[\frac{2}{1 + e^{-b\frac{\pi}{\omega}}\left(\sin\pi\frac{\omega_0}{\omega}\,\mathrm{tg}\,\varphi + \cos\pi\frac{\omega_0}{\omega}\right)} - 1\right], \tag{20}$$

where

$$\sin\alpha = \frac{\omega_0}{\sqrt{\omega_0^2 + b^2}}, \quad \cos\alpha = \frac{b}{\sqrt{\omega_0^2 + b^2}}, \quad \mathrm{tg}\,\varphi = \frac{\sin\alpha - e^{-b\frac{\pi}{\omega}}\sin\left(\pi\frac{\omega_0}{\omega} - \alpha\right)}{\cos\alpha - e^{-b\frac{\pi}{\omega}}\cos\left(\pi\frac{\omega_0}{\omega} - \alpha\right)}.$$

Figure 3b, gives the characteristic in (20) for small damping ($b \ll \omega_0$). For a proper choice of feedback parameters (w_0 somewhat smaller than ω_{p_1}), one may obtain a larger autooscillation frequency than with ordinary lagged feedback. Since the characteristic in (20) does not intersect the axis of abscissas, the feedback increases the autooscillation frequency for all values of the load.

A Feedback Contactless Relay

Increasing the autooscillation frequency of a system with a VR has an adverse effect on the operating conditions of the relay contacts, which lowers the reliability of the regulator. Therefore, in vibration regulators of electric motor speeds, there are sometimes used contactless relays, constructed on the basis of linear amplifiers with overcompensated feedback (magnetic amplifiers in relay modes, electronic relays, etc.). The resetting ratio of these relays depends on the magnitude of the positive feedback. The frequency of speed autooscillation is the lower, the higher the relay's resetting ratio [3]. Therefore, in designing the relay, the tendency is to decrease the magnitude of the feedback. However, a significant decrease in the size of the feedback leads to an increase in the lag time and to instability of the relay characteristic. Therefore, it is necessary to find a compromise solution in computing the positive feedback.

A contactless relay may have a significant lag time (particularly in magnetic contactless relays), which decreases the autooscillation frequency in a system with VR. The use of a rigid, or of a lagged, negative feedback for increasing the autooscillation frequency is inadmissible when using a contactless relay of this type, since this latter decreases the effectiveness of the positive feedback. When contactless relays are used, one may quite successfully employ flexible feedback paths. Flexible feedback is widely used to decrease the lag time of magnetic contactless relays [5-7]. Decreasing the relay's lag increases the autooscillation frequency in a system with VR. We now give an example of choosing the parameters of a flexible feedback.

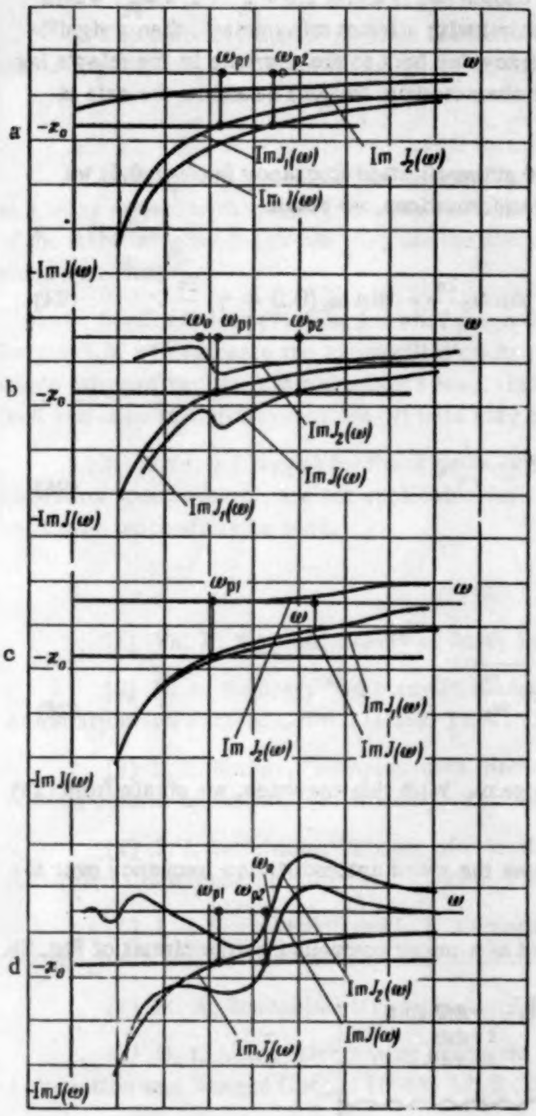


Fig. 3

As will be shown below, by means of a flexible feedback it is possible to increase the autooscillation frequency, not only due to a decrease in the relay's lag, but also due to a change in the regulator parameters.

As an example, we shall demonstrate the method of choosing the flexible feedback parameters for the case, considered above, of a relay system for controlling the speed of an asynchronous motor using contactless control apparatus – a magnetic amplifier and a magnetic contactless relay [3] (Fig. 2b). In this case, the relay's lag time significantly affects the autooscillation frequency. Taking into account of the dependence of relay lag on the current in the flexible feedback circuit leads to quite complicated expressions. Therefore, the analysis given below attempts to show only the qualitative features of the influence of the feedback parameters on the autooscillation frequency. The rough quantitative relationship resulting from this analysis must be made more precise empirically.

Let the flexible feedback element have the transfer function $W(p) = Cp/(LCp^2 + Crp + 1)$ (circuit in Fig. 2b). The characteristic $\mathrm{Im}J_2(\omega)$ depends on the form of the characteristic equation's roots. If the roots are real ($r/2 > \sqrt{L/C}$, $p_{1,2} = -b \pm a$), then the characteristic $\mathrm{Im}J_2(\omega)$ has the form (in this case $x = +z$):

$$\mathrm{Im}\, J_2(\omega) = \frac{U}{2aL} \frac{e^{p_1 \frac{\pi}{\omega}} - e^{p_2 \frac{\pi}{\omega}}}{1 + e^{p_2 \frac{\pi}{\omega}} + e^{p_1 \frac{\pi}{\omega}} + e^{-b \frac{2\pi}{\omega}}}. \tag{21}$$

The characteristic in (21) for $L \ll C$ (Fig. 3c) shows that the flexible feedback varies the system parameters in such fashion that the autooscillation frequency is decreased for $\omega_{p_1} > \omega_1$. On the other hand, the flexible feedback decreases the relay's lag [5] and thereby increases the autooscillation frequency. Consequently, it is necessary so to choose the flexible feedback parameters that $\mathrm{Im}J_2(\omega) \approx 0$ (Fig. 3c).

If the roots are complex ($r/2 < \sqrt{L/C}$, $p_{1,2} = -b \pm j\omega_0$) then, after some transformations, we obtain

$$\mathrm{Im}\, J_2(\omega) = \frac{U e^{-b\frac{\pi}{\omega}} \sin \pi \frac{\omega_0}{\omega}}{\omega_0 L \left(1 + e^{-b\frac{\pi}{\omega}}\, 2\cos \pi \frac{\omega_0}{\omega} + e^{-2b\frac{\pi}{\omega}}\right)}, \tag{22}$$

$$\mathrm{Re}\, J_2(\omega) = \frac{U e^{-b\frac{\pi}{\omega}} \left(\cos \pi \frac{\omega_0}{\omega} - \frac{b}{\omega_0} \sin \pi \frac{\omega_0}{\omega} + e^{-b\frac{\pi}{\omega}} \cos \pi \frac{\omega_0}{\omega} + 2\right)}{\omega L \left(1 + e^{-b\frac{\pi}{\omega}}\, 2\cos \pi \frac{\omega_0}{\omega} + e^{-2b\frac{\pi}{\omega}}\right)^2}. \tag{23}$$

Figure 3d, gives the characteristic in (22) for $b \ll \omega_0$.

If the parameters of the flexible feedback are properly chosen (ω_0 is about 1.3 ω_{p_1} or 1.4 ω_{p_1}, where ω_{p_1} is the autooscillation frequency in a system without feedback but with a lesser relay delay), then a significant increase may be obtained in the speed autooscillation frequency due both to the decrease in the relay's lag and to the changes in the regulator parameters. In this case, the characteristic $\mathrm{Im}J_2(\omega)$ intersects the axis of abscissas.

In order to elucidate how the feedback affects the speed autooscillation frequency for $\gamma \pm 0.5$, we equate the second term in expression (8b) to zero. After some transformations, we obtain

$$e^{-b\frac{2\pi}{\omega}} \sin \omega_0 (0{,}5-\gamma)\frac{2\pi}{\omega} - e^{-b(0{,}5-\gamma)\frac{2\pi}{\omega}} \sin \omega_0 \frac{2\pi}{\omega} = \sin \omega_0 (0{,}5+\gamma)\frac{2\pi}{\omega}. \qquad (24)$$

It is easily understood that condition (24) holds if

$$\frac{\omega_0}{\omega} = \frac{m}{2} \ ; \ \gamma = 0{,}5 \pm \frac{l}{m}, \qquad (25)$$

where m and l are arbitrary positive integers.

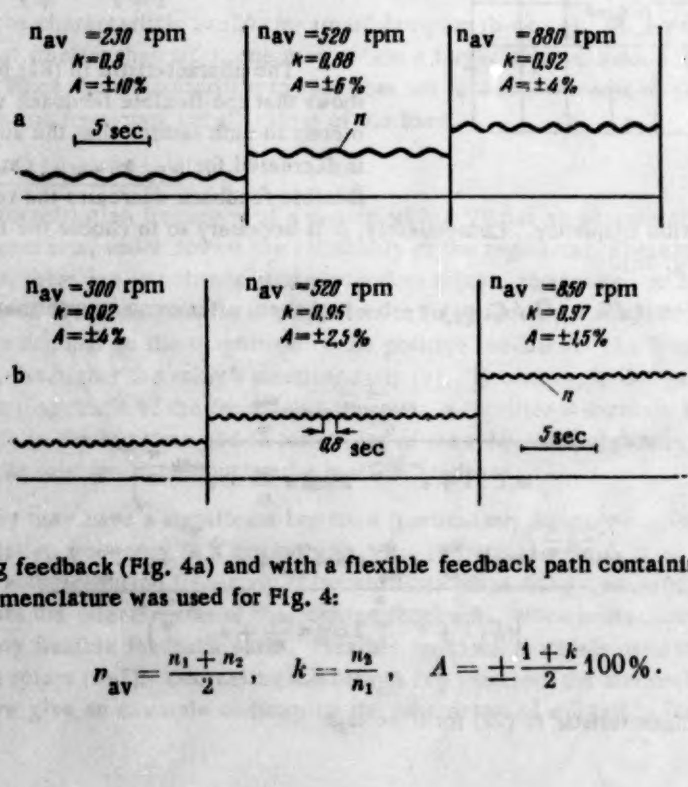
By taking (13) into account, we obtain

$$l = \sqrt{0.25\, m^2 - \frac{\omega_0}{\omega_{p_1}}}. \qquad (26)$$

It follows from (26) that l can be an integer only for large m. With this condition, we obtain from (25) and (26) that $\gamma_1 \approx 1$ and $\gamma_2 \approx 0$.

Thus, the feedback parameters chosen for $\gamma = 0.5$ increase the speed autooscillation frequency over almost the entire range of load variation.

Figure 4 shows the oscillograms of the variations in speed of a motor controlled by the circuit of Fig. 2b,



both without correcting feedback (Fig. 4a) and with a flexible feedback path containing an oscillatory link (Fig. 4b). The following nomenclature was used for Fig. 4:

$$n_{av} = \frac{n_1 + n_2}{2}, \quad k = \frac{n_2}{n_1}, \quad A = \pm \frac{1+k}{2} 100\%.$$

The correctly chosen flexible feedback containing an oscillatory link can give the very best effect of increasing the autooscillation frequency in a system with contactless relays.

SUMMARY

1. A system of electric motor speed control with vibration regulators, for $\gamma = 0.5$, can always be reduced to a relay system with symmetric relay characteristics and no dead zone. In this case, for choosing the parameters of the stabilizing feedback one may use the simplest construction employed for relay systems with symmetric relay characteristics.

2. Conditions (7) show for what range of load variation the feedback, whose parameters were chosen for $\gamma = 0.5$, will increase the autooscillation frequency. In the majority of cases, conditions (7) hold over the whole range of variation of the motor's load. In the case when $\mathrm{Im}J_2(\omega) = 0$ for some values of ω, the range of load variation in which conditions (7) hold may be found from expressions (8).

3. Rigid and lagged feedback paths, which are ordinarily used for stabilizing vibration systems of electric motor speed control, are not applicable for certain types of contactless relays. In such cases, flexible feedback may successfully be used.

LITERATURE CITED

[1] Ya. Z. Tsypkin, Theory of Relay Systems for Automatic Control. [in Russian]. Gostekhizdat (1955).

[2] N. A. Korolev, "On periodic modes in relay systems with internal feedback loops," [in Russian]. Automation and Remote Control (USSR) 17, 11 (1956).

[3] L. L. Rotkov, "Sampled-data methods for regulating asychronous electric motor speeds using contactless control apparatus," [in Russian]. Vestnik Élektropromyshlennosti 1 (1958).

[4] S. I. Bernshtein, "Theory of vibration regulators of electrical machines," [in Russian]. Trydi II All-Union Conference on Automatic Control Theory, Vol. 1. Izdatelstvo AN SSSR (1955).

[5] I. B. Negnevitskii and L. L. Samurina, "Experimental investigation of transient responses in contactless magnetic relays," [in Russian]. Élektrichestvo 9 (1953) .

[6] M. A. Rozenblat, Magnetic Amplifiers. [in Russian]. Published by "Sovetskoe Radio" (1956).

[7] O. I. Aben, "Decreasing magnetic amplifier lag by introducing flexible feedback," [in Russian]. Automation and Remote Control (USSR) 18, 2 (1957).

Received February 16, 1959

ON THE TRANSFER FUNCTION OF AN ASYNCHRONOUS TWO-PHASE MOTOR

V. G. Kutvinov

(Podol'sk)

The paper considers the special features inherent in the determination of the coefficients of the transfer function of a motor for amplitude and phase modulation. An engineering method for determining the transfer function is presented. An analysis is carried out of the effect of nonlinearity in the motor control characteristics.

Basic Relationships

The electromagnetic processes in a two-phase motor which determine its rotational torque and speed are described by the system of equations [1]

$$E_c = T_c R_c + \frac{d\Psi_c}{dt}, \quad E_e = R_e I_e + \frac{d\Psi_e}{dt},$$
$$0 = i_c r + \frac{d\Psi_c}{dt} - \omega_m \phi_B, \quad 0 = i_e r + \frac{d\Psi_e}{dt} + \omega_e \phi_e,$$
$$T = MI_e i_e - MI_e i_c,$$

where T is the motor's rotational torque, ω_m is the nominal rotational speed of the rotor; the subscript "c" denotes elements of the control circuit and currents in this circuit, the subscript "e" as the same use for the excitation circuit. The elements of the motor's power supply circuit are shown in Fig. 1 (the excitation circuit has analogous elements and is not shown), in accordance with which we use the following notation in the sytem of equations in (1):

$$K_0 U_g = E_c, \quad R_c = R_e + R_1, \quad \Psi_c = I_c (L_{s1} + M), \quad \phi_y = (L_{s2} + M) i_c.$$

If we write the voltages and currents in complex form, using the theorem on equivalent generators, the system of equations in (1) can be generalized for any motor supply circuit, including the circuit of Fig. 2:

$$\dot{A}_c \dot{E}_y = \dot{Z}_c \dot{I}_c + \dot{Z}_m (\dot{I}_c - i_c), \quad 0 = i_c \dot{Z}_2 + \dot{Z}_m (i_c - I_c) + \\ + j \frac{\omega_m}{\omega_0} [\dot{Z}_m \dot{I}_e - (\dot{Z}_m + \dot{Z}_2 - r) i_e], \tag{2}$$

$$\dot{A}_e\dot{E}_e = \dot{Z}_e\dot{I}_e + \dot{Z}_m(\dot{I}_m - \dot{i}_e), \quad 0 = \dot{i}_e\dot{Z}_2 + \dot{Z}_m(\dot{i}_e - \dot{I}_e) -$$

$$- j\frac{\omega_m}{\omega_0}[\dot{Z}_m\dot{I}_y - (\dot{Z}_m + \dot{Z}_2 - r)\,\dot{i}_c], \tag{2}$$

$$T = \mathrm{Re}\,[M\dot{I}_c\overset{*}{i}_e - M\dot{I}_e\overset{*}{i}_c].$$

Here, we use the notation: $\dot{Z}_m = j\omega_0 M$, $\dot{Z}_2 = r + j\omega_0 L_{s2}$,

$$\dot{Z}_c = \dot{Z}_{ic} + R_1 + j\omega_0 L_{s1}, \quad \dot{Z} = \dot{Z}_{ie} + R + jL_s\omega_0,$$

$\dot{Z}_{ic}$, $\dot{Z}_{ie}$, $\dot{A}_c\dot{E}_c$ and $\dot{A}_e\dot{E}_e$ are the internal impedances and emf's of the equivalent generators of the control and excitation circuits,respectively; the sign * here and in the sequel denotes the complex conjugate of a quantity.

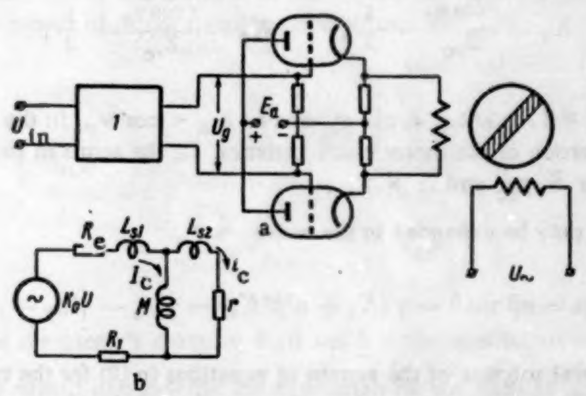


Fig. 1. a) is the circuit, without transformers, of the motor connections, 1 is the modulator and ac amplifier, b is the replacement circuit of the output stage, K_0 is the gain of this stage, R_e is the output impedance of the stage, L_{s1} and R_1 are the induction leakage and active impedance of the motor's stator winding, L_{s2} and r are the same for the rotor winding and M is the coefficient of mutual induction of the stator and rotor windings.

We now introduce the relative variables $\mu = T/T_s$ and $\gamma = \omega_m/\omega_r$.

Here, T_s is the motor's starting torque for a symmetric supply ($\alpha\beta = \sin\theta = 1$):

$$T_s = 2\frac{E^2 p_d}{\omega_0 Z_{ke}}\frac{r}{Z_{re}}\left|1 - \frac{\dot{Z}_{re}}{\dot{Z}_{xe}}\right|\frac{10^4}{1.02}\ \mathrm{g\cdot cm} \tag{3}$$

$\omega_r = \omega_0/p_d$ is the synchronous motor speed, p_d is the number of pairs of poles of the stator winding, $\alpha e^{j\varphi} = \dot{E}_c/\dot{E}_e$ is the symmetry coefficient of the supplying voltages, $\beta e^{j\varphi_1} = \dot{Z}_{ke}/\dot{Z}_{kc}$ is the symmetry coefficient symmetry coefficient of the motor supply circuit, $\dot{Z}_{kc}$ and $\dot{Z}_{ke}$ are the transmission impedances of the control and excitation circuits with the rotor braked, $\theta = \varphi + \varphi_1$ is the phase difference of the currents in the motor's stator windings with the rotor braked, Z_{re} is the modulus of the excitation circuit impedance, measured on the rotor's side, $Z_{re}e^{j\varphi_e} = \dot{Z}_2 + \dot{Z}_e\dot{Z}_m/(\dot{Z}_e + \dot{Z}_m)$, E is the actual value of emf of the supply source for the excitation circuit.

Solution of the system of equations in (2) for the constant-speed steady state leads to the expression

$$\mu = \alpha\beta\sin\theta - \frac{\gamma(K_1+\alpha^2\beta^2K_2) - \gamma^2\alpha\beta\sin\theta\left[4K_1K_2 - \frac{r^2\cos(\varphi_c-\varphi_e)}{Z_{rc}Z_{re}}\right]}{1-\gamma^2\left[1+4K_1K_2-2\frac{r^2\cos(\varphi_y-\varphi_n)}{Z_{rc}Z_{re}}\right]+\gamma^4\left(2K_1-\frac{r^2}{Z_{rc}^2}\right)\left(2K_2-\frac{r^2}{Z_{re}^2}\right)} -$$

$$- \frac{\gamma^3\left[\frac{r^2}{2Z_{rc}^2}-K_1+\alpha^2\beta^2\left(\frac{r^2}{2Z_{re}^2}-K_2\right)\right]+\gamma^4\left(2K_1-\frac{r^2}{Z_{rc}^2}\right)\left(2K_2-\frac{r^2}{Z_{re}^2}\right)}{1-\gamma^2\left[1+4K_1K_2-2\frac{r^4\cos(\varphi_y-\varphi\ )}{Z_{rc}Z_{re}}\right]+\gamma^4\left(2K_1-\frac{r^2}{Z_{rc}^2}\right)\left(2K_2-\frac{r^2}{Z_{re}^2}\right)}. \tag{4}$$

Here, $Z_{rc}e^{j\varphi_c}$ is the complex impedance of the impedance of the control circuit, measured on the rotor's side. The coefficients K_1 and K_2 are defined by the expressions

$$K_1 = \frac{r\cos\varphi_c}{Z_{rc}} - \frac{1}{2}, \quad K_2 = \frac{r\cos\varphi_e}{Z_{re}} - \frac{1}{2}. \tag{5}$$

Since $|K_1| \leq 1/2$, $|K_2| \leq 1/2$, $r/Z_{rc} < \cos\varphi_c$ and $r/Z_{re} < \cos\varphi_e$, in the limit, $0 \leq \gamma \leq 0.8$, i.e., at the limits of the operating portion of the motor characteristics, all the terms in the denominator of expression (4) are less than unity, if $2r \gtrless Z_{rc}$ and $2r \gtrless Z_{re}$.

Therefore, expression (4) may be expanded in the series

$$\mu = \alpha\beta\sin\theta - \gamma(K_1+\alpha^2\beta^2K_2) - \gamma^2 n_2 - \gamma^3 n_3 - \ldots \tag{6}$$

It is obvious that the general solution of the system of equations in (2) for the transient response can also be given as the sum of an infinite series in powers of γ. But $\gamma < 1$, and the changes in speed, $\Delta\gamma$, are much less than unity, this being the more valid, the higher the frequency of speed change. Based on this, one may substitute in the general solution, with a high degree of accuracy, $\gamma\Delta\gamma \approx \Delta\gamma^2 \approx \Delta\gamma^3 \approx - 0$.

Then the general solution for the transient response of the system of equations in (2) leads to an expression which differs from expression (4) only in the coefficients of the zero and first powers of γ. It is therefore now possible, in solving the system of equations in (2), to limits oneself to the first power of γ and to seek the solution by setting the average speed equal to zero.

Let the input of the amplitude modulator be acted upon by a harmonic function of frequency Ω. Then, the voltage in the motor's control circuit will be

$$e_c = K_N\frac{E_m}{2}\cos[(\omega_0-\Omega)t+\varphi_N] + K_e\frac{E_m}{2}\cos[(\omega_0+\Omega)t+\varphi_e], \tag{7}$$

where $\dot{K}_N = K_N e^{j\varphi_N}$ and $\dot{K}_e = K_e e^{j\varphi_e}$ are the complex transmission factors of the amplifier and modulator for the frequencies of $\omega_0\ \Omega$ and $\omega_0 + \Omega$, and E_m is the amplitude of the signal voltage at the modulator input.

Further, let the following voltage act on the excitation circuit

$$e_e = E_0\cos(\omega_0 t+\varphi_3). \tag{8}$$

The solution of the system of equations in (2) for these voltages leads to an expression whose real part defines the rotational torque of the motor as a function of its speed and controlling voltage:

$$\mu(p) = \alpha\frac{\beta K_0}{4}\operatorname{Re} e^{pt}\left[\frac{\dot{K}_N^*}{\dot{K}_0^*}\frac{\frac{s}{\omega_0}(\dot{Z}_2+\dot{Z}_m)e^{-j\theta}}{r\left(\frac{\dot{Z}_{sr}}{\dot{Z}_{rc}}\right)^*\left(\frac{\dot{Z}_s+sM}{\dot{Z}_c+\dot{Z}_m}\right)^*} - j\frac{\dot{K}_e}{\dot{K}_0}\frac{(\dot{Z}_{qq}+qM)e^{j\theta}}{r\left(\frac{\dot{Z}_{rq}}{\dot{Z}_{rc}}\right)\left(\frac{\dot{Z}_q+qM}{\dot{Z}_c+\dot{Z}_m}\right)}\right] -$$

$$-\frac{\gamma}{4}\,\mathrm{Re}\,e^{pt}\left[\frac{\dot{Z}_m\,qM}{\dot{Z}_{qr}\,(\dot{Z}_q+qM)}+\frac{\dot{Z}_2+\dot{Z}_m}{(\dot{Z}_{sr})^*}\right]+\mu^*. \tag{9}$$

Here, μ^* is the multinomial which is the complex conjugate, with respect to p, of the first two terms of the right member, $P = j\Omega$, $s = j\omega_0 - j\Omega$, $q = j\omega_0 + j\Omega$, $\dot{K}_0 = Ke^{j\varphi_0}$ is the complex transmission factor of the amplifier and modulator at frequency ω_0, $\dot{Z}_s$ and $\dot{Z}_q$ are the impedances of the stator winding control supply circuit at the frequencies, respectively, of $\omega_0 - \Omega$ and $\omega_0 + \Omega$, $\dot{Z}_{sr}$ and $\dot{Z}_{qr}$ are impedances of the motor control circuit, measured on the rotor's side, for the frequencies, respectively, of $\omega_0 - \Omega$ and $\omega_0 + \Omega$ and $\dot{Z}_{2q} = r + j(\omega_0 + \Omega)L_{s2}$.

The torque μ (p) is balanced by the moment of inertia and the frictional torque. If the dry friction torque is ignored, the variable component of motor speed may be written as

$$\mu(p) = \gamma\left(a + P\frac{\tau\omega_0}{p_d}\right), \tag{10}$$

where

$$\tau = \frac{J}{T_s}, \quad a = \frac{k\omega_0}{T_s\,p_d}, \tag{11}$$

J is the moment of inertia of the motor's charging shaft and k is the coefficient of viscous friction.

Equations (9) and (10) permit one to write the expression for the transfer function

$$W(P) = \frac{\omega_0}{E_0 p_d}\,\frac{\beta K_0 \sin\theta\, A(P)}{\left(a + P\frac{\tau\omega_0}{p_d}\right)D(P) + K_1 N(P)},$$

in which the operators A(P), D(P) and N(P) are defined by the expressions

$$A(P) = \mathrm{Re}\,\frac{1}{2}\left[\frac{\dot{K}_N^* e^{j\theta}(\dot{Z}_2+\dot{Z}_m)}{\dot{K}_0 \sin\theta}\,\frac{s}{\omega_0}\,\frac{\dot{Z}_{qr}}{\dot{Z}_{2C}}\,\frac{\dot{Z}_q+qM}{\dot{Z}_C+\dot{Z}_m} - j\,\frac{\dot{K}_e}{\dot{K}_0}\,\frac{e^{-j\theta}}{\sin\theta}\,(\dot{Z}_{2q}+qM)\left(\frac{\dot{Z}_s+sM}{\dot{Z}_C+\dot{Z}_m}\right)^*\left(\frac{\dot{Z}_{sr}}{\dot{Z}_{rC}}\right)^*\right], \tag{13a}$$

$$D(P) = \left(\frac{\dot{Z}_{sr}}{\dot{Z}_{rC}}\right)^*\frac{\dot{Z}_{qr}}{\dot{Z}_{rC}}\left(\frac{\dot{Z}_s+sM}{\dot{Z}_C+\dot{Z}_m}\right)^*\frac{\dot{Z}_q+qM}{\dot{Z}_C+\dot{Z}_m}, \tag{13b}$$

$$N(P) = \mathrm{Re}\,\frac{1}{2K_1}\left[\frac{\dot{Z}_m^2}{\dot{Z}_{2C}(\dot{Z}_C+\dot{Z}_m)}\,\frac{q}{j\omega_0}\left(\frac{\dot{Z}_{sr}}{\dot{Z}_{rC}}\right)^*\left(\frac{\dot{Z}_s+sM}{\dot{Z}_C+\dot{Z}_m}\right)^* + \frac{\dot{Z}_2+\dot{Z}_m}{\dot{Z}_{rC}^*}\,\frac{\dot{Z}_{qr}}{\dot{Z}_{rC}}\left(\frac{\dot{Z}_s+sM}{\dot{Z}_C-\dot{Z}_m}\right)^*\frac{\dot{Z}_q+qM}{\dot{Z}_C+\dot{Z}_m}\right], \tag{13c}$$

and the symbol "Re" defines the real coefficient with respect to P.

The expression just obtained for the transfer function is difficult to use in practical computations, since direct calculation of the coefficients by means of expressions (13) necessitates the carrying out of rather lengthy mathematical transformations.

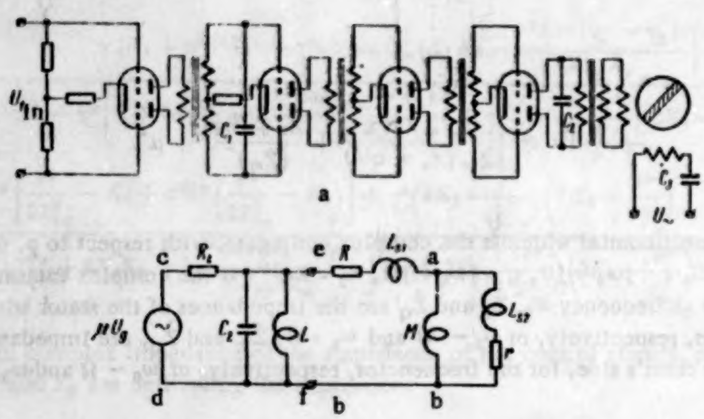


Fig. 2. a) is the circuit for a four-stage amplifier with a two-phase motor at its output, b) is the replacement circuit of the amplifier's output stage, "e-f" are the input terminals of the motor winding, μ and R_i are the gain and internal impedance of the output tube, U_g is the voltage signal at the tube's control grid, L is the induction of the transformer's output winding, C_1 = 0.1 microfarad, C_2 = 0.2 microfarad and C_3 = 6 microfarad.

However, the method for determining the transfer function can be essentially simplified if, in expression (12), the operator N(P) is replaced by D(P), i.e., if one writes

$$W(P)=\frac{\omega_0 K_0}{p_g E_0}\frac{\beta \sin 0}{a+K_1+P\dfrac{\tau\omega_0}{p_g}}\frac{A(P)}{D(P)}. \tag{14}$$

With this simplification, the coefficients of the transfer function will be determined with a certain error, the relative size of which depends on the relationships of the coefficients of $\tau\omega_0/p_d$.

For the simplest motor connection circuit (Fig. 1), the following values of the coefficients of operators D(P) and N(P) (with the assumption that induction leakage in the motor winding equalled zero) were obtained from expressions (13).

The coefficients were computed for a motor with R_1 = 20 ohms, r = ω_0M = 50 ohms.

For more complicated motor connection circuits, the computation of the coefficients ofoperators D(P) and N(P) involves rather difficult transformations of expressions (13) which are impossible to carry out without the aid of computers. However, it may be assumed that even in the case of more complex circuits, the coefficients of D(P) and N(P) do not differ from each other more than the amounts given in the foregoing table since, as follows from expressions (13), they are defined by the absolute values of the circuit impedances.

As shown by a comparison of the data on existing two-phase motors, the values of the coefficients of $\tau\,\omega_0^2/p_d$ are usually not less than ten to fifteen, which exceeds the coefficients of operators N(P) and D(P) by a factor of ten to twenty. With these relationships, the replacement of operator N(P) by D(P) leads to an error in determining the transfer function coefficients which does not exceed 5% to 7%, which is not beyond the limits of engineering calculation accuracy.

If motor control is implemented by varying the initial phase of the controlling voltage

$$e_c'=E_m\cos(\omega_0 t+\Delta\varphi\sin\Omega t),$$

then, for the voltage in the motor control circuit, in the limit of small changes in the initial phase, one may write, instead of expression (7),

$$e_c = E_m K_0 \cos \omega_0 t - \frac{\Delta\varphi E_m}{2} K_N \cos[(\omega_0 - \Omega)t + \varphi_N] + + \frac{\Delta\varphi E_m}{2} K_e \cos[(\omega_0 + \Omega)t + \varphi_e]. \tag{15b}$$

In this case, the average speed equals zero for $\theta = 0$. Therefore, by repeating all the mathematical transformations, and making the substitution $\theta_f = \theta - \pi/2$, we obtain, instead of (12), the following expression for the transfer function:

$$W(P) = \frac{\omega_m(P)}{\varphi(P)} = \frac{\omega_0}{p_d} \frac{\alpha\beta \sin\theta_f A(P)}{\left(a + P\frac{\tau\omega_0}{p_d}\right) D(P) + K_1 N(P)}. \tag{16}$$

After replacing operator N(P) by D(P), we get

$$W(P) = \frac{\omega_0}{p_d} \frac{\alpha\beta \sin\theta_f}{a + K_1 + p\frac{\tau\omega_0}{p_d}} \frac{A(P)}{D(P)}. \tag{17}$$

Expressions (14) and (17) allow one to avoid the lengthy mathematical transformations connected with the determination of the transfer function coefficients. The operator A(P)/D(P) is determined from equations (2) for the short-circuited mode, and it is therefore possible to employ a graphical method which is not only simpler, but more revealing.

Graphical Determination of the Function A(P)/D(P)

Let the controlling voltage defined by expression (7) act on the magnetizing loop of the motor's control winding, and the voltage defined by (8) on the magnetizing loop of the motor's excitation winding (points a and b on Fig. 2b). Then, the currents in the motor's stator winding will be

$$\begin{aligned} I_c &= -\frac{E_m K_e \sin[(\omega_0 + \Omega)t + \varphi_e]}{2M(\omega_0 + \Omega)} - \frac{E_m K_N \sin[(\omega_0 - \Omega)t + \varphi_N]}{2M(\omega_0 - \Omega)}, \\ I_e &= -\frac{E \sin(\omega_0 t + \varphi_3)}{M\omega_0}. \end{aligned} \tag{18}$$

We determine the current in the motor's rotor windings with the condition that the induction loss in the rotor windings is ignored:

$$\begin{aligned} i_c &= \frac{E_m K_e \cos[(\omega_0 + \Omega)t + \varphi_e]}{2r} + \frac{E_m K_N \cos[(\omega_0 - \Omega)t + \varphi_N]}{2r}, \\ i_e &= \frac{E\cos(\omega_0 t + \varphi_3)}{r}. \end{aligned}$$

This simplification cannot lead to significant error, since the coefficients of operator A(P)/D(P) are defined by the frequency characteristics of the multiloop circuit (modulator, amplifier and motor – Fig. 2a) and the relative influence of the induction leakage from the motor's rotor windings is negligibly small.

If we substitute the values of current from expressions (18) and (19) in the torque equation (1), we get

$$T = T_0\alpha_0 \sin(\varphi_3 - \varphi_0)\left[\frac{K_N(2\omega_0 - \Omega)}{4K_0(\omega_0 - \Omega)} \cos(\Omega t - \varphi_N + \varphi_0) - - \frac{K_e(2\omega_0 + \Omega)}{4K_0(\omega_0 + \Omega)} \cos(\Omega t + \varphi_e - \varphi_0)\right] + T_0\alpha_0 \cos(\varphi_3 - \varphi_0) \times \tag{20}$$

$$\times\left[\frac{K_N(2\omega_0-\Omega)}{4K_0(\omega_0-\Omega)}\sin(\Omega t-\varphi_N+\varphi_0)-\frac{K_e(2\omega_0+\Omega)}{4K_0(\omega_0+\Omega)}\sin(\Omega t+\varphi_e-\varphi_0)\right]. \quad (20)$$

Here,

$$\alpha_0=\frac{K_0E_m}{E_0}, \quad T_0=\frac{E_0^2}{\omega_0 r}.$$

By bearing in mind that $\sin\Omega t=\cos(\Omega t-\pi/2)$, we have

$$T_\sim=T_0\alpha_0\sin(\varphi_3-\varphi_0)\left[\frac{\dot K_N^*(2\omega_0-\Omega)}{4\dot K_0^*(\omega_0-\Omega)}+\frac{\dot K_{..}(2\omega_0+\Omega)}{4\dot K_0(\omega_0+\Omega)}\right]-$$
$$-jT_0\alpha_0\cos(\varphi_3-\varphi_0)\left[\frac{\dot K_N^*(2\omega_0-\Omega)}{4\dot K_0^*(\omega_0-\Omega)}-\frac{\dot K(2\omega_0+\Omega)}{4\dot K_0(\omega_0+\Omega)}\right]. \quad (21)$$

If motor control is implemented by varying the initial phase of the controlling voltage then, by assuming that a voltage defined by expression (15b) acts on the magnetizing loop of the motor's control circuit, we obtain, after mathematical transformations similar to (18)-(20).

$$T=T_0\alpha_0\sin(\varphi_3-\varphi_0)-\Delta\varphi T_0\alpha_0\sin(\varphi_3-\varphi_0)\times$$
$$\times\left[\frac{K_N(2\omega_0-\Omega)}{4K_0(\omega_0-\Omega)}\cos(\Omega t-\varphi_N+\varphi_0)-\frac{K_e(2\omega_0+\Omega)}{4K_0(\omega_0+\Omega)}\cos(\Omega t+\varphi_e-\varphi_0)\right]-$$
$$-\Delta\varphi T_0\alpha_0\cos(\varphi_3-\varphi_0)\left[\frac{K_N(2\omega_0-\Omega)}{4K_0(\omega_0-\Omega)}\sin(\Omega t-\varphi_N+\varphi_0)+\right. \quad (22)$$
$$\left.+\frac{K_e(2\omega_0+\Omega)}{4K_0(\omega_0+\Omega)}\sin(\Omega t+\varphi_e-\varphi_0)\right].$$

We note further that, with phase modulation, $\alpha\beta=\text{const}\approx 1$, $-\pi/2\le\varphi_3-\varphi_0\le\pi/2$ with the average speed equal to zero for $\varphi_3-\varphi_0=0$, while for amplitude modulation $(\varphi_3-\varphi_0)=\text{const}\approx\pi/2$. We therefore carry out the substitution $\sin(\varphi_3-\varphi_0)=\cos(\varphi_3-\varphi_0-\pi/2)$. In addition, we introduce the substitution $\cos\Omega t=\sin(\Omega t-\pi/2)$ since, in the given case, the input modulating function varies sinusoidally (15a). For the variable component of the motor torque, we write the following equation in complex quantities:

$$T_\sim=\Delta\varphi T_0\alpha_0\left\{\sin\left(\varphi_3-\varphi_0-\frac{\pi}{2}\right)\left[\frac{\dot K_N^*(2\omega_0-\Omega)}{4\dot K_0^*(\omega_0-\Omega)}+\frac{\dot K_e(2\omega_0+\Omega)}{4\dot K_0(\omega_0+\Omega)}\right]-\right.$$
$$\left.-j\cos\left(\varphi_3-\varphi_0-\frac{\pi}{2}\right)\left[\frac{\dot K_N^*(2\omega_0-\Omega)}{4\dot K_0^*(\omega_0-\Omega)}-\frac{\dot K_e(2\omega_0+\Omega)}{4\dot K_0(\omega_0+\Omega)}\right]\right\}. \quad (23)$$

The torque $T_0\alpha_0\sin(\varphi_3-\varphi_0)$ can be expressed in terms of the voltage acting on the input terminals (points c and d on Fig. 2b) of the motor control circuit, starting from the condition of torque equality:

$$T_0\alpha_0\sin(\varphi_3-\varphi_0)=T_s\alpha\beta\sin\theta,\quad \Delta\varphi T_0\alpha_0\sin\left(\varphi_3-\varphi_0-\frac{\pi}{2}\right)=T_s\Delta\varphi\alpha\beta\sin\theta_f.$$

Instead of expressions (21) and (23), we write

$$\mu=\alpha\beta\sin\theta\left\{\frac{\dot K^{*}_{\mathrm{N}}(2\omega_0-\Omega)}{4\dot K^{*}_{0}(\omega_0-\Omega)}+\frac{\dot K_{\mathrm{e}}(2\omega_0+\Omega)}{4\dot K_0(\omega_0+\Omega)}-\right.$$
$$\left.-j\frac{\cos\theta}{\sin\theta}\left[\frac{\dot K^{*}_{\mathrm{N}}(2\omega_0-\Omega)}{4\dot K^{*}_{0}(\omega_0-\Omega)}-\frac{\dot K_{\mathrm{e}}(2\omega_0+\Omega)}{4\dot K_0(\omega_0+\Omega)}\right]\right\},\tag{25a}$$

$$\mu_{\mathrm{f}}=\Delta\varphi\alpha\beta\sin\theta_{\mathrm{f}}\left\{\frac{\dot K^{*}_{\mathrm{N}}(2\omega_0-\Omega)}{4\dot K^{*}_{0}(\omega_0-\Omega)}+\frac{\dot K_{\mathrm{e}}(2\omega_0+\Omega)}{4\dot K_0(\omega_0+\Omega)}-\right.$$
$$\left.-j\frac{\cos\theta_{\mathrm{f}}}{\sin\theta_{\mathrm{f}}}\left[\frac{\dot K^{*}_{\mathrm{N}}(2\omega_0-\Omega)}{4\dot K^{*}_{0}(\omega_0-\Omega)}-\frac{\dot K_{\mathrm{e}}(2\omega_0+\Omega)}{4\dot K_0(\omega_0+\Omega)}\right]\right\}.\tag{25b}$$

It follows from expressions (25) that, in both cases, the function A(P)/ D(P) can be written in the form

$$\frac{A(P)}{D(P)}=\frac{\dot K^{*}_{\mathrm{N}}}{4\dot K^{*}_{0}}\,\frac{2\omega_0-\Omega}{\omega_0-\Omega}+\frac{\dot K_{\mathrm{e}}}{4\dot K_0}\,\frac{2\omega_0+\Omega}{\omega_0+\Omega}-$$
$$-j\frac{\cos\theta}{\sin\theta}\left[\frac{\dot K^{*}_{\mathrm{N}}}{4\dot K^{*}_{0}}\,\frac{2\omega_0-\Omega}{\omega_0-\Omega}-\frac{\dot K_{\mathrm{e}}}{4\dot K_0}\,\frac{2\omega_0+\Omega}{\omega_0+\Omega}\right].\tag{26}$$

The most essential difference between the transfer functions for phase and amplitude modulation is that, with amplitude modulation, the angle θ does not vary, while for phase modulation the angle θ_{f} varies as a function of the value of the input modulating function. Correspondingly, the factor $\cos\theta_{\mathrm{f}}/\sin\theta_{\mathrm{f}}$ in expression (26) also varies and, consequently, so does the function A(P)/ D(P).

For amplitude modulation, the angle θ = const and, in practice, equals $\pi/2$, therefore we may write, instead of expression (26), the simpler:

$$\frac{A(P)}{D(P)}=\frac{\dot K^{*}_{\mathrm{N}}(2\omega_0-\Omega)}{4\dot K^{*}_{0}(\omega_0-\Omega)}+\frac{\dot K_{\mathrm{e}}(2\omega_0+\Omega)}{4\dot K_0(\omega_0+\Omega)}.$$

It is impossible to make a similar simplification for the case of phase modulation. Expressions (26) and (27) allow the amplitude-phase characteristics of the corresponding functions A(P)/ D(P) to be found by graphical constructions.

In the majority of cases, the electric circuit from the modulator to the motor consists of several noninterconnected contours. The complex transmission factors of these contours as functions of frequency, K(jω), may be given for some simplification of the environment. For example, for the typical contour circuits shown in Fig. 3, it suffices to ignore the transformer's induction leakage and the capacitance of the mounting.

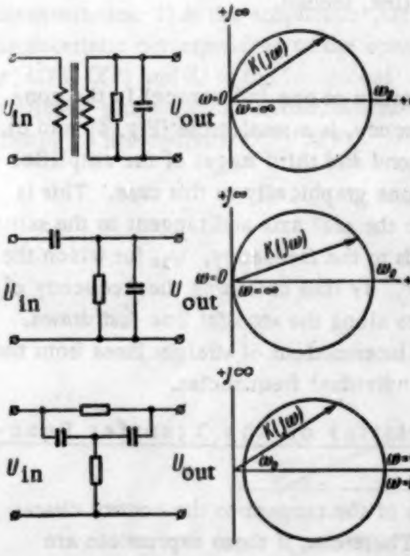


Fig. 3. Circuits of typical contours and their frequency characteristics.

The complex transmission factor as a function of frequency may be obtained for the whole circuit by simply multiplying the complex transmission factors of the individual contours which, again, is conveniently done graphically.

Figure 4 gives the frequency characteristics, K(jω), for the individual stages of the given concrete amplifier circuit of Fig. 2. The frequency characteristic of the entire circuit of Fig. 2. The frequency characteristic of the entire circuit and the amplitude-phase characteristic corresponding to the operator A(P)/ D(P) are given in Fig. 5.

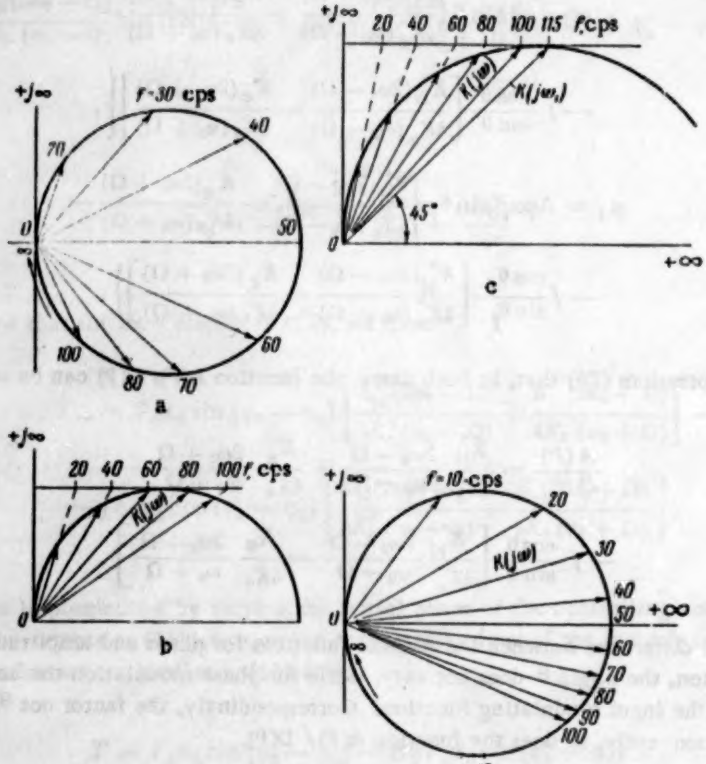


Fig. 4. Frequency characteristics of the amplifier stages. a, b, c, and d are the characteristics, respectively, of the first, second, third and fourth stages.

In those cases when there is only one reactive element (one capacitance or one inductance) in the contour, the complex transmission factor of the contour as a function of frequency, is a semicircle (Fig. 4b and c). To this type of contour belong, for example, the anode circuits of the second and third stages of the amplifier shown in Fig. 2. The laying out of the frequencies is also conveniently done graphically in this case. This is done in the following way (Fig. 4c). If we draw a straight line parallel to the real axis and tangent to the semicircle, then the point of intersection of this line and the circle corresponds to the frequency, ω_1, for which the argument of the complex transmission factor equals 45°: arg $(K(j\omega_1))$ =.45°. By thus obtaining the frequency of the point of intersection, we obtain the scale for laying out the frequencies along the straight line just drawn. The dispolsition of the frequencies around the circle is obtained from the intersections of straight lines from the origin of coordinates to the points on the scale line corresponding to the individual frequencies.

Effect of Nonlinearity in the Motor Control Characteristics on the Transfer Function

The coefficient $a + K_1$ in expressions (14) and (17) defines the slope of the tangent to the control characteristic $\gamma = f(\alpha)$ on the initial portion of this characteristic (for $\gamma \to 0$). Therefore, if these expressions are extended to the case of an arbitrary value of average motor speed, it is necessary to substitute for $a + K_1$ in expressions (14) and (17) some coefficient $K(\gamma)$ which defines the tangent to the control characteristics at any arbitrary point.

By substituting for μ in expression (3) its value from the relationship $\mu = a\gamma$, and then differentiating with respect to a and γ, one can show that, for any practically possible cases of viscous frictional load on the motor, and for any motor parameters, the value of $K(\gamma)$ can vary between the limits $-0.5 \leq K(\gamma) \leq 16$ as the average motor speed varies from $\leq = 0$ to $\leq = 0.8$.

Figure 6 shows the reciprocal amplitude-phase characteristics corresponsponding to the circuit of Fig. 2 for various values of the coefficient $K(\gamma)$ within the limits stated.

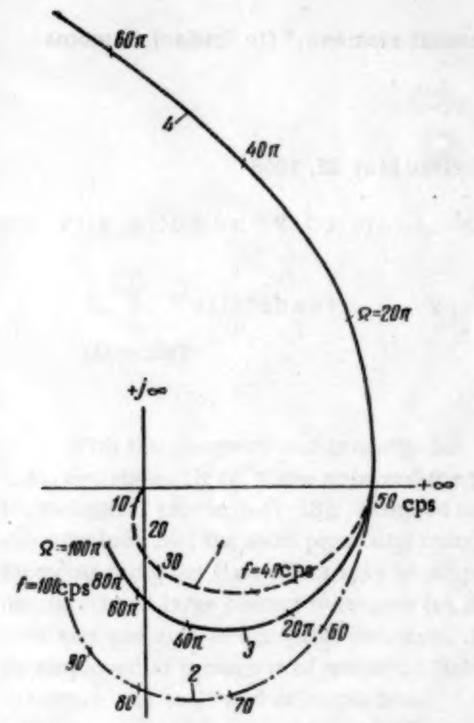


Fig. 5. Frequency characteristic of the entire circuit. 1) is the conjugate low-frequency arm of the circuit characteristic, 2) is the high-frequency arm of the circuit characteristics; 3) is the amplitude-phase characteristic corresponding to the operator A(P)/ D(P) and 4) is the reciprocal amplitude-phase characteristic, corresponding to the operator D(P)/ A(P).

As may be concluded from Fig. 6, the worse conditions for stability of motor speed control correspond to an average speed equal to zero.

It is advantageous, in practical computations, to orient oneself by these worst conditions and, therefore, in these computations, it should be assumed that K(γ) equals K_1 + a. In this case, the coefficients of the transfer function are defined by relationships (4), (5) and (11). The complex ratios, r/Z_{rc}, and r/Z_{re}, necessary for substitution in the formulae listed, are determined by experiment on the motor when free-running, short-circuited and synchronously opposed.

SUMMARY

1. The transfer functions for the ac circuits with two-phase motors differ, for amplitude (14) and phase (17) modulation, only by constant coefficients but, with phase modulation, the transfer function is nonlinear due to the dependence of the operator A(p)/D(p) on the value of the input function.

2. When there is negative feedback of the speed, stable regulation is possible for any value of motor parameters if the feedback gain is chosen with the condition

$$K_\omega \geqslant \frac{p_d E_0}{2K_\theta \omega_0 \beta \sin\theta}.$$

3. The use of a graphical method of determining the transfer function of the modulator-demodulator link allows, with sufficient simplicity, the analysis to be carried out of the effect of nonlinearity of the motor's transfer function even when the motor is operating with other, including nonlinear elements.

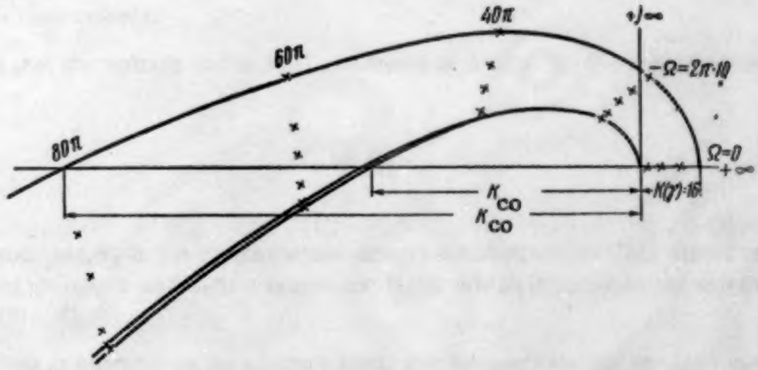


Fig. 6. Reciprocal amplitude-phase characteristic of the link amplifier two-phase-motor for various values of the coefficient K(γ). 1) is for K(γ) = −0.5; 2) is for K(γ) = 0; 3) is for K(γ) = 16.

LITERATURE CITED

[1] I. M. Sadovskii, "Asynchronous electric motors as control circuit elements," [in Russian]. Automation and Remote Control (USSR) 13, 6 (1952).

Received May 12, 1958

ON THE ACCURACY OF HALL ELEMENTS

L. S. Vasil'chenko, L. V. Sentyurina and B. S. Sotskov

(Moscow)

With the discovery and investigation of semiconducting materials with large drift mobilities, such as InSb, InAs, Ge, HgSe, HgTe, there appeared the practical possibility of using the Hall effect for solving a series of technological problems [1-13]. Based on the characteristics of these materials, as shown in Table 1 [1], one can conclude that the most promising materials, from the point of view of practical usage, are InSb and Ge. Elements using the Hall effect may be employed as magnetic field strength measurers, magnetic induction measurers and large current measurers (up to a thousand amperes); they may be used as transducers, modulators, detectors and also as computer elements. Elements using a change in ohmic resistance in magnetic fields can be employed as measurers of magnetic fields, magnetic induction and current, as transducers, modulators and detectors, and as dc and ac amplifiers.

We shall consider the application of elements using the Hall effect (Hall elements) for the measurement of magnetic flux and we shall show, based on an investigation of the possible errors, the expediency of constructing such an instrument.

We first give certain basic functional relationships for Hall elements.

When a flat conductor is placed in a magnetic field (Fig. 1), there is a tilt of the equipotential surfaces by an angle θ which can be found from the expression

$$\tan\ \theta = \frac{3\pi}{8}\mu_n B = \frac{E_y}{E_x},$$

where $\mu_n = V_e$ E is the electron mobility in cm^2/second volt, B is the magnetic induction in volt·second/cm^2, V_e is the speed of electron motion in cm/second and E_x and E_y are the electric field strengths, in volt/cm, along the x and y axis, respectively.

It is well known that the voltage at the Hall electrodes is defined by the relationship

$$U_y = R_H \frac{I_x B}{d} C_r,$$

where R_H is the Hall constant, I_x is the current whose direction coincides with the direction of the x axis, d is the thickness of the flat specimen and C_r is a correction factor which depends on the relationship between the specimen dimensions (Fig. 2).

When a Hall element operates on an external load, the voltage between the Hall electrodes, and the current in the circuit of these electrodes, are defined by the expressions

$$U_y = r_{in}(B) I_y + U_{yL}, \quad I_y = \frac{U_{yL}}{r_L},$$

TABLE 1

Material	Specific impedance ρ, in ohms/centimeter	Mobility, μ, in centimeter^3 per volts^{-1}seconds^{-1}	Mobility μ, in centimeters^2 per volts^{-1}seconds^{-1}		Hall constant, R_H, in centimeters^{3}coulomb^{-1}	Efficiency η, for H = 10^3 oersted	Transmission factor k, for H = 10^3 oersted	$\Delta\rho/\rho$, in %; for H = 10^4 oersted	Voltage sensitivity γ in microvolts per oersted
			electrons	holes					
Copper	$1,6\cdot10^{-6}$	—	27	—	$-5\cdot10^{-3}$	$2\cdot10^{-10}$	—	—	$2\cdot10^{-10}$
Bismuth	$1\cdot10^{-4}$	—	$5\cdot10^{3}$	—	-6	0,09	—	0,1	$3\cdot10^{-4}$
Silicon	$6,36\cdot10^{4}$	—	$1.4\cdot10^{3}$	350	-10^{8}	<0,1	—	0,006	—
Germanium	50	10^{14}	$3.6\cdot10^{3}$	1800	-10^{5}	0,05	0,045	0.05	540
Cuprous oxide	$>10^{4}$	—	—	100	10^{6}	<0,1	—	—	—
Indium stibide	$1\cdot10^{-2}$	10^{16}	$60\cdot10^{3}$	700	-500	12,4	0.71	16	210
Indium arsenide	$2,5\cdot10^{-1}$	10^{15}	$30\cdot10^{3}$	200	-910^{2}	3,0	0,35	3,4	500
Mercury selenide	$1\cdot10^{-3}$	$5\cdot10^{17}$	$10\cdot10^{3}$	—	—	0,32	0,06	0,06	40
Mercury telluride	$6,7\cdot10^{-3}$	$5\cdot10^{18}$	$10\cdot10^{3}$	~500	—	—	0,12	0,4	10

where $r_{in\,(B=0)}$ is the element's internal impedance between the Hall electrodes in the absence of a magnetic field, U_{yL} is that portion of the Hall voltage which drops across the load, r_L, and r_L is the load impedance.

We denote by λ the ratio of r_L to $r_{in(B=0)}$. Then,

$$U_y = U_{yL}\left(\frac{1}{\lambda} + 1\right).$$

The relationship between the Hall voltage and the magnetic induction is not linear, since with increasing B, both U_y and r_{in} increase. The deviation from linearity will be the less, the lower, for a given specimen, is $\Delta\rho/\rho = f(B)$, where ρ is the specific impedance of the elements material, and $\Delta\rho$ is the absolute change in specimen impedance when the latter is placed in a magnetic field.

The sensitivity of the element to the voltage between the Hall electrodes, as a function of the magnitude of the magnetic induction, is determined from the formula

$$\gamma = \frac{U_{y\,\max}}{B} = \frac{R_H I_{x\max}}{d}\, 10^{-8}, \text{volts/oersted}$$

By expressing I_x in terms of the admissible power leakage, $P_{em.adm}$, we get

$$\gamma = b\sqrt{\frac{2\mu_n R_H P_{em\cdot adm}}{ld}}\cdot 10^{-8}, \text{ volts/oersted}$$

Figure 3 shows the relationship $U_y = f(B, \lambda)$. Ordinarily, values of $\lambda = r_L\ r_{in(B=0)}$ are chosen for which there is a linear relationship between U_y and B. This magnitude are λ, designated by λ_{opt}, is one of the parameters of a Hall-element. Usually, even for λ_{opt}, there is a certain deviation from linearity in the relationship $U_y = f(B,\lambda)$, as is shown in Fig. 4. We denote the greatest value of this deviation by ϵ_{max}, and find its value for λ (Fig. 5). The least value of ϵ_{max} will correspond to λ_{opt}.

The error when a Hall element is used to measure magnetic fields may be estimated in the following way. When using an external (to be measured) magnetic field lying between the limits of B = 0 and B = 10^4 gauss, a certain nonlinearity occurs in the relationship $U_y = f(B, \lambda)$. The minimum value of the error, ΔU_y^* corresponds to some value of λ, which must be determined from the graph of Fig. 5.

Moreover, there must be taken into account the residual voltage U'_{yo} which, for existing methods of preparing Hall elements, lies within the limits of $(0.25 \text{ to } 2.0) \cdot 10^{-3}$ volts. To obtain an instrument with a given error (i.e., in a given class of accuracy) N, it is necessary that the upper limit of the voltage taken off from the Hall electrodes equals

$$U_{y\,\max} = \frac{U'_{y0} + \Delta U'_{y}}{N}.$$

By bearing in mind that $\Delta U'_y = \epsilon_{\max} U_{y\,\max}$, we get

$$U_{y\,\max} = \frac{U'_{y0}}{N - \epsilon_{\max}}.$$

The magnitude of the current in the circuit supplying electrons is determined from

$$I = \frac{U_{y\,\max} d}{C_r R_H B_{\max}}.$$

It is necessary that this value of current be less than I_{adm}, defined by the condition

$$I_{adm} = b \sqrt{\frac{\mu_t d}{0.24\rho}\,\theta'_{adm}}$$

where b is the width of the plate, d is its thickness, μ_t is the total heat transfer coefficient, θ'_{adm} is the admissible superheating temperature, $\theta'_{adm} = \theta_{\max} - \theta_0$, where θ_0 is the ambient temperature; for Ge, $\theta_{\max} =$ = 65 degrees C.

If ρ depends essentially on the temperature, then the following formula should be used

$$I_{adm} = b \sqrt{\frac{\mu_t d t'_{adm}}{0.24\rho_0 (1 + t'_{adm})}}.$$

For existing units, the following values are characteristic: $\epsilon_{\max} = 0.5\%$ to 1.0%, $I_{adm} = 0.075$ to 0.6 amperes, $U_{y\,\max} = 0.045$ to 0.450 volts.

In case the Hall element is used in magnetic systems with ferromagnetic materials, account should be taken of the additional error due to the magnetization hysteresis characteristic. Figure 6a, shows the magnetization cycle of a ferromagnetic material and, on Fig. 6b, there is shown the resulting magnetization characteristic of a magnetic system with an airgap equal to δ. The slope of the cycle axis equals

$$\alpha = \operatorname{arctg}\left(\frac{S_a}{\delta}\,\frac{l_m}{S_m}\,\frac{n_H}{n_B}\right),$$

where S_a is the cross-section of the airgap, l_m and S_m are the length and cross section of the magnetic circuit, n_B and n_H are the scales of the ordinate and abscissa axes respectively. If the coercivity equals H_c, then the greatest error in magnetic induction will equal $2\Delta B = 2S_a l_m H_c / \delta S_m$, which corresponds to the additional error from the voltage

$$\Delta U_y \leqslant \frac{2C_r R_H I_{adm} \Delta B}{d}.$$

Thus, the necessary value of $U_{y\,max}$ will equal

$$U_{y\,max} = \frac{U'_{y0} + \Delta U'_y + \Delta U''_y}{N}.$$

The accuracy class attained is

$$N = \frac{U'_{y0} + \Delta U''_y}{U_{y\,max}} + \varepsilon_{max}$$

Finally, it is necessary to take into account the errors arising from the effect of temperature. This effect manifests itself in two ways. First, the Hall coefficient varies with changes in temperature. In the limits of $\pm 20°$ C to 100° C, this change can usually be considered to equal

$$R_H = R_{H_0}[1 + \alpha_t(\theta_H - 20°)],$$

where R_{H_0} is the Hall coefficient at the temperature 20° C, α_t is the temperature coefficient of the Hall coefficient variation; α_t = 0.0150/degree for InSb, = 0.005/degree for InAs and = −0.0002/degree for mixed crystals.

Fig. 1. A Hall element in a magnetic field.

Moreover, the value of r_{in} changes in accordance with the law, $r_{in} = r_{in,0}[1 + \alpha'_t(\theta_H - 20°)]$, where $r_{in,0}$ is the internal impedance of the specimen at 20° C, α'_t is the temperature coefficient of the impedance variation, θ_H is the working temperature. For example, for InAS $\alpha'_t = -0.0023$/degree. A change in r_{in} leads to a change in the value of λ and, consequently, increases ε_{max}, the nonlinearity error. To take the temperature errors into account, it is necessary to determine the possible changes in r_{in} and λ, and to find the new value of the nonlinearity error ε_{max} corresponding to the value of λ. Moreover, the error from the variation of the Hall coefficient should be found, this being equal to

$$\Delta U''' = R_H R_{H20} \frac{B_{max} I_{adm}}{d} \alpha_t(\theta_H - 20°).$$

The final error of the instrument is defined by the expression

$$N = \frac{U'_{y0} + \Delta U'_y + \Delta U'''_y}{U_{y\,max}} + \varepsilon'_{max}.$$

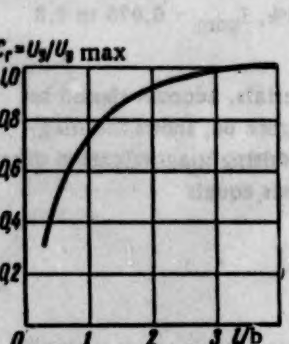


Fig. 2. Relative value of Hall voltage as a function of the ratio of specimen length to width, l/b.

We now determine the instrument error for a Hall element prepared of indium arsenide (InAs), with the following conditions.

1. The maximum working temperature is +60° C.
2. The nominal value of the magnetic field equals 10,000 gauss.
3. The magnetic systems of the instrument have one and the same form and dimensions, but are prepared from three different materials: Mo-Permalloy, electrical steel É3A and transformer steel É4A.

The construction of the magnetic system is shown schematically in Fig. 7. The thickness of the magnetic circuit block is 1 cm., δ = 0.05 cm., S_a = 1.6 cm², l_m = 12.6 cm, S_m = 1.6 cm².

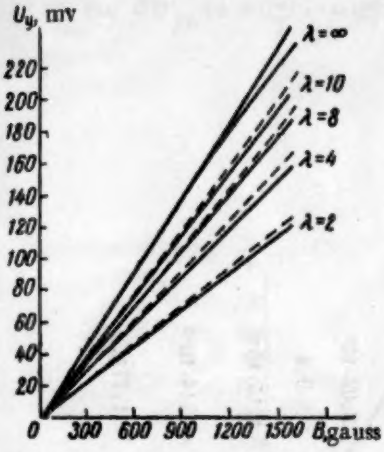


Fig. 3. The Hall voltage as a function of the magnetic field induction for various values of the parameter λ for a Hall-element of germanium.

The parameters of indium arsenide Hall elements, and a number of data necessary for computing instrument error, are given in Table 2.

The results of the error determinations for three ferromagnetic materials used in the preparation of magnetic circuits are given in Table 3.

In connection with the attempts to use germanium for the preparation of Hall elements, we determine the maximum error of such instruments, starting from experimental data. Figure 8 gives the dependence of the voltage between the Hall electrodes on the magnetic fields existing between the limits of $+B_{max}$ and $-B_{max}$. It is clear from Fig. 8 that U_y has the form of a hysteresis characteristic, and that the error due to hysteresis, ΔU_y^*, about 11 millivolts.

The error due to temperature variation, ΔU_y^{**}, according to Fig. 9, is 123 millivolts, assuming that $U_{y\ max}$ = 200 millivolts and $\theta_H - \theta_0$ = 25° C.

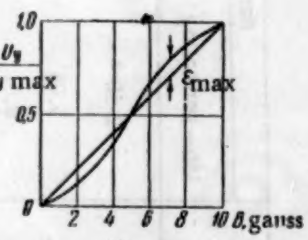


Fig. 4

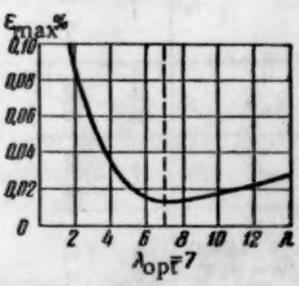


Fig. 5

Fig. 4. The Hall voltages as a function of the magnetic field induction for the case of alinear dependence of U_y on B, and for an actual case.

Fig. 5. The coefficient ε as a function of the ration of load impedance to Hall-element impedance.

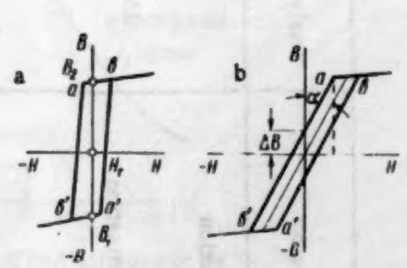


Fig. 6

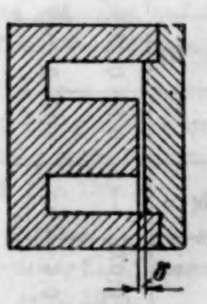


Fig. 7

By using Fig. 3, we may determine ε_{max}, assuming that the resistance of the specimen decreases by a factor of about 1.5 when the temperature changes from 25° C:

$$\varepsilon'_{max} = \frac{U_y - U_{y\,lin}}{U_{y\,lin}} \approx 0{,}025.$$

TABLE 2

	For temperature 20° C												For temperature 60° C		
	I_{adm}, amp	B, kg	δ, mm	U_y, mv	r_{L} r_{in} $(B=0)$ for linear λ_{opt}	ε, %	R_H, cm^3 · k^{-1}	r_{in} $(B=0)$, ohm	c_r	U, v	α_t, 1/	α_t', per °C	λ_{opt}', %	ε, %	r_{in}' $(B=0)$, ohm
Minimal value	0.075	—	0.5	0.25	—	0.5	—	0.5	—	0.045	—	—	—	—	—
Maximal value	0.6	10	1.0	2.00	7	1.0	100	1.0	0.95	0.45	$-5 \cdot 10^{-4}$	$-2.3 \cdot 10^{-[illegible]}$	7.7	1.3	0.91

TABLE 3

Source of error	Analytical expression for error	Ferromagnetic material		
		Mo – permalloy	electrical steel	transformer steel
Nonlinearity in function $U_y = f(B)$	$\Delta U_y' = \varepsilon_{max} U_{y\,max}$	$2 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$2 \cdot 10^{-3}$
Hysteresis of magnetization characteristic	$2\Delta B = 2S_b l_m H_c / \delta S_m$, gauss	5.04	$2.77 \cdot 10^{2}$	$2.02 \cdot 10^{2}$
	H_c, oersted	0.01	0,55	0,4
	$\Delta U_y'' = 2\,\dfrac{C_r R\ I_{adm} \Delta B}{d}$	$2.87 \cdot 10^{-4}$	$1.58 \cdot 10^{-2}$	$1.15 \cdot 10^{-2}$
Temperature effect	$\Delta U_y''' = \dfrac{C_r R_H B_{max} I_{adm}}{d} \alpha_t (t - 20°)$	$-1.14 \cdot 10^{-3}$	$-1.14 \cdot 10^{-3}$	$-1.14 \cdot 10^{-3}$
Summed instrument error	$N = \dfrac{U_y' + \Delta U_y'' + \Delta U_y'''}{U_{y\,max}} + \varepsilon'$	1.28%	2.72%	1.77%

For $\Delta U'_{yo} \ll \Delta U''_y + \Delta U'''_y$ and $U_{y\,max} = 200$ millivolts, we have

$$N_{Ge} = \frac{\Delta U''_y + \Delta U'''_y}{U_{y\,max}} + \varepsilon_{max} = 0.692.$$

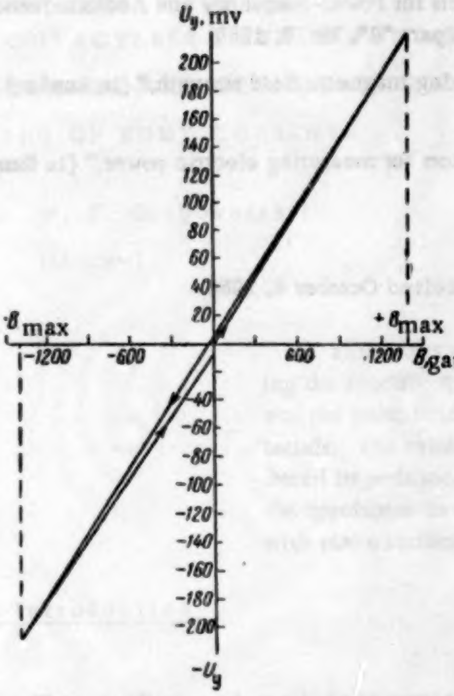


Fig. 8. Hysteresis dependency of Hall voltage on magnetic field induction for a Germanium-N5. Hall-element. l = 0.82 cm., b = 0.42 cm., d = 0.032 cm, R_H = 1335 ohms, I_x = 6 milliamperes.

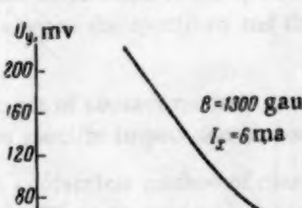


Fig. 9. The Hall voltage as a function of temperature for a germanium Hall element for B = 1300 gauss and I_x = 6 milliamperes.

It is thus clear that in the case of a germanium Hall element, the instrument error is determined on the basis of the error due to variation of the Hall voltage with temperature. Consequently, to increase the accuracy class of such an instrument, temperature compensation is necessary. The problem of temperature compensation for temperature variations of ± 5 10 ± 10° is solvable. For a wide temperature range (Δt = 60° C), the problem is significantly complicated. In designing instruments for measuring magnetic flux based on the Hall effect, it is advantageous to use semiconducting materials with low temperature dependencies of the Hall voltage, for example, InAs. In this case, the total instrument error does not exceed 2% on the average.

LITERATURE CITED

[1] V. P. Zhuze and A. R. Regel'. Technical Use of the Hall Effect. [in Russian]. Published in Leningrad by the Dom Nauchno-Tekhnichesk. Propagandy, vol. 11 (1957).

[2] V. S. Sotskov. Elements of Automatic and Remote Control Apparatus. [in Russian]. Gosénergoizdat (1952).

[3] A. R. Regel'. Semiconductor Measurers of Magnetic Field Strength. [in Russian]. Dom Nauchno-Tekhn. Propagandy (Leningrad, 1956).

[4] O. D. Elpat'evskaya and A. R. Regel', "On certain capabilities of measuring magnetic field strength by surface-layer Hall elements prepared from HgSe and HgTe, and their solid solutions," [in Russian]. J. Tech. Phys. (USSR), 26 (1956) p. 2432.

[5] E. M. Barlow and L. M. Stephenson. The Hall Effect and Its Application to Power Measurement at Microwave Frequencies, Proc. I. E. E. (part B), v. 103, 167, 1956.

[6] J. M. Ross and N. A. C. Thompson, An Amplifier Based on the Hall Effect. Nature, v. 175, No. 4455, 1955.

[7] "Application of Hall generators in modern measuring technology," [in German]. Deutsche Elektrotechnik 10, 2 (1956), p. 10.

[8] V. N. Bogomolov, "DC amplifiers with transformers based on the effect of changes in semiconductor impedance in a magnetic field," [in Russian]. J. Tech. Phys. (USSR), 26 (1956) p. 2480.

[9] V. N. Bogomolov, "Certain new types of instruments built of semiconductors (New uses of the Hall effect)," [in Russian]. J. Tech. Phys. (USSR) 26 (1956) p. 693.

[10] E. W. Saker, F. A. Cunnell and J. T. Edmond, Indium Antimonide as a Fluxmeter Material, Brit. Journ. Appl. Phys. v. 6, N. 6, pp. 217-220, 1955.

[11] H. E. M. Barlow, The Design of Semiconductor Wattmeters for Power-Frequency and Audiofrequency Applications. Proceedings of the Institute of Electr. Engin., v. 102, part "B", No. 2, 1955.

[12] G. E. Pikus and O. V. Sorkin, "A new method for measuring magnetic field strength," [in Russian]. J. Tech. Phys. (USSR) 27 (1957) pp. 26, 47.

[13] L. S. Berman, "The use of the Hall effect in semiconductors for measuring electric power," [in Russian]. J. Tech. Phys. 27 (1957), pp. 1192, 1197.

Received October 6, 1958

A CONTACTLESS METHOD FOR MEASURING SPECIFIC IMPEDANCE AND GEOMETRIC DIMENSIONS BY MEANS OF EDDY CURRENTS

V. P. Grabovetskii

(Moscow)

This paper considers a contactless method for measuring the electric specific impedance of nonmagnetic materials and the geometric dimensions of large specimens of these materials. The relationships are provided which relate the introduced impedances and the parameters of the specimens, when the specimens have the form of cyclinders, prisms and blocks with plane surfaces.

I. Introduction

The specific impedance is an important characteristic of materials, and its value is, in itself, of significant interest. Moreover, from the magnitude of the specific impedance one can predict many physical characteristics of the material. Thus, a knowledge of the specific impedance renders it possible to control the constituents and quality of a material. However, until now this possibility has hardly been exploited, since the existing contact methods of measuring specific impedance have a number of disadvantages. To make measurements, it is necessary to prepared special specimens. In measuring specimens of short length and large cross-sections, the impedance of the specimen becomes commensurable with the unstable transient impedances of the contacts between the specimen and the measuring circuit, thanks to which the accuracy of measurement is decreased.

The use of contact methods makes it difficult to implement continuous automatic control of the magnitude of the specific impedance on continous production lines.

The contactless method of measuring specific impedance by means of eddy currents is free of these disadvantages. The principle underlying this method amounts to the following. The specimen to be inspected is placed in the variable high-frequency magnetic field of a coil. Eddy currents are induced in the specimen, concentrated in its surface layers. These currents induce their own magnetic flux, which is directed toward the basic flux. The resulting magnetic flux of the coil is varied, which engenders a change in the coil's active and reactive impedances. The degree of this change depends on the specimen's specific impedance, its geometric dimensions, its form and on any defects present in it. Consequently, from the variations in the coil's parameters one can judge, not only the magnitude of the specific impedance, but also the specimen's geometric dimensions: diameter, lamina thickness, plating thickness, etc. Moreover, the method of eddy currents can be used to discover blow holes, bubbles, cracks and structural inhomogeneities in the specimen, to mention a few of the possible flaws. Thus, the eddy current measuring method has broad possibilities, and is very promising for use in automatic inspection systems.

A number of works [1-4 and others] are devoted to the eddy current method. Förster [3] obtained the

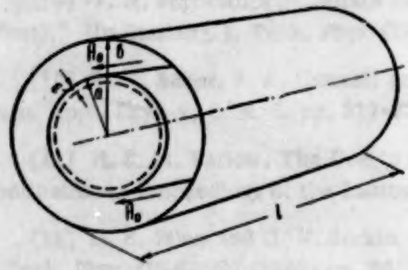


Fig. 1

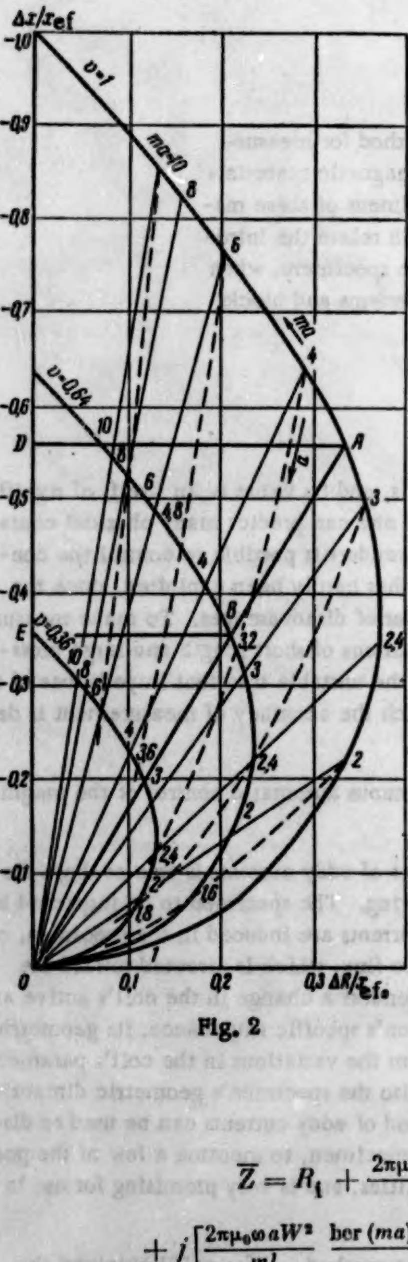


Fig. 2

relationship connecting the magnitude of the voltage induced in the measuring coil with the parameters of the cylindrical specimens being tested, where the source of the electromagnetic field was an exciting coil coaxial with the measuring coil and the specimens. Förster introduced the concept of the specimen's "effective magnetic permeability," by means of which one takes into account the nonuniformity of the magnetic field distribution across the specimen's cross-section due to the surface effect phenomenon. This concept distorts the physical nexus of the phenomenon, and leads the author to erroneous conclusions. What is very laudable in Förster's work is his use of the complex plane for representing the relationships connecting the voltage on the measuring coil with the specimen's parameters. This gave him the capability of measuring the specific impedance of the specimen independently of variations in its geometric dimensions and, conversely, of measuring the specimen's geometric dimensions independently of variations in its specific impedance. These results of Förster are used by us.

In this paper, the approach to the solution of the problem is somewhat different; we consider the relationships between the impedances introduced in the coil and the specimen parameters when the coil is simultaneously exciting and measuring. The knowledge of such dependencies is necessary when the coil-transducer is connected to circuits which react to variations in the coil impedances.

II. Inspection of Cylindrical Specimens

We now consider how the parameters of a solenoid vary when a cylindrical bar of nonmagnetic material is introduced into it (Fig. 1). We shall assume that the lengths of the solenoid and bar are sufficiently great that we may consider the conditions of the problem to be homogeneous in the direction of the solenoid axis, and we shall further assume that the frequency ω of the sinusoidally varying control flowing through the coil is sufficiently high for surface effects to ensue.

For this we adopt an equivalent solenoid circuit, consisting of a series connection of active and inductive impedances. If the measurements are made at frequencies less than 1/10 the solenoid's resonant frequency, ω_0, then we may, with an error not greater than 1% , ignore the shunting effect of the solenoid's natural capacitance, and consider the active and inductive impedances of the equivalent circuit to be the actual values of impedance. All further conclusions are valid for frequencies less than 0.1 ω_0.

The expression for the total solenoid impedance when a long cylindrical specimen is introduced into it was obtained by Dwight [5]:

$$\overline{Z} = R_1 + \frac{2\pi\mu_0\omega a W^2}{ml} \frac{\operatorname{ber}(ma)\operatorname{ber}'(ma) + \operatorname{bei}(ma)\operatorname{bei}'(ma)}{\operatorname{ber}^2(ma) + \operatorname{bei}^2(ma)} + j\left[\frac{2\pi\mu_0\omega a W^2}{ml} \frac{\operatorname{ber}(ma)\operatorname{bei}'(ma) - \operatorname{bei}(ma)\operatorname{ber}'(ma)}{\operatorname{ber}^2(ma) + \operatorname{bei}^2(ma)} + \frac{\pi\mu_0\omega W^2}{l}(b^2 - a^2)\right]. \tag{1}$$

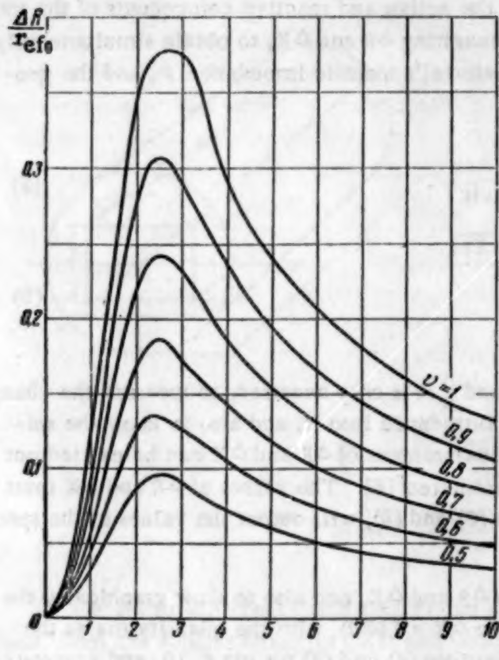


Fig. 3

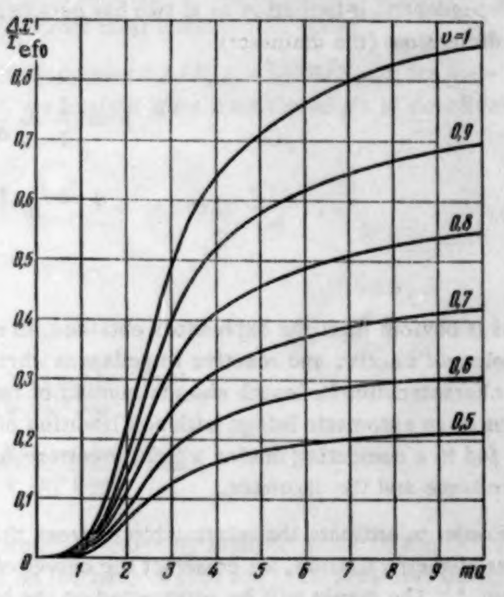


Fig. 4

Here, R_l is the solenoid's inherent active impedance, $\mu_0 = 0.4\pi \cdot 10^{-8}$ ohm second/cm is the magnetic permeability, W is the number of turns in the solenoid, a is the specimen's radius, b is the solenoid's radius, l is the length of the solenoid and the specimen, $m = \sqrt{\mu_0 \omega \gamma}$, γ is the specimen's specific conductivity, ber and bei are, respectively, the real and imaginary components of the Bessel function $I_0(ma\sqrt{j})$, and ber' and bei' are, respectively, and derivatives of ber and bei with respect to ma.

If we separate the real and imaginary parts of expression (1) and subtract from them the corresponding values of active and reactive impedance in the solenoid with no bar present (i.e., with a = 0)

$$R_{\mathrm{ef}\,0} = R_l, \tag{2}$$

$$X_{\mathrm{ef}\,0} = \frac{\pi\mu_0\omega W^2 b^2}{l}, \tag{3}$$

we obtain the values of the impedances introduced into the solenoid loop due to the introduction into it of the cylindrical bar of nonmagnetic material:

$$\Delta R = \frac{2\pi\mu_0\omega W^2 a}{ml}\,\frac{\mathrm{ber}\,(ma)\,\mathrm{ber}'\,(ma) + \mathrm{bei}\,(ma)\,\mathrm{bei}'\,(ma)}{\mathrm{ber}^2\,(ma) + \mathrm{bei}^2\,(ma)}, \tag{4}$$

$$\Delta X = \frac{2\pi\mu_0\omega W^2 a}{ml}\,\frac{\mathrm{ber}\,(ma)\,\mathrm{bei}'\,(ma) - \mathrm{bei}\,(ma)\,\mathrm{ber}'\,(ma)}{\mathrm{ber}^2\,(ma) + \mathrm{bei}^2\,(ma)} - \frac{\pi\mu_0\omega W^2 a^2}{l}. \tag{5}$$

For values of ma > 10, we can use the asymptotic expressions for the Bessel function, and transform expressions (4) and (5) to the form

$$\Delta R \approx \frac{\pi\mu_0\omega W^2 a}{l}\sqrt{\frac{2}{\mu_0\omega\gamma}}, \tag{6}$$

$$\Delta X \approx \frac{\pi\mu_0\omega W^2 a}{l}\sqrt{\frac{2}{\mu_0\omega\gamma}} - \frac{\pi\mu_0\omega W^2 a^2}{l}. \tag{7}$$

The introduction of the bar into the solenoid changes both the active and reactive components of the total solenoid impedance. This circumstance makes it possible, by measuring ΔR and ΔX, to obtain simultaneously, and independently, information as to two bar parameters – the material's specific impedance ρ, and the geometric dimensions (the diameter):

$$\rho = \frac{1}{\gamma} = \frac{\overline{\Delta R}^2}{\Delta R - \Delta X} \frac{l}{2\pi W^2}, \tag{8}$$

$$a = \frac{1}{W} \sqrt{\frac{(\Delta R - \Delta X)\, l}{\pi \mu_0 \omega}}. \tag{9}$$

As is obvious from the expressions obtained, to measure ρ and a it is only necessary to measure the changes in the solenoid's active and reactive impedances when the bar is introduced into it, and also to know the solenoid's characteristics its length and the number of turns in it. Measurement of ΔR and ΔX can be carried out by means of an automatic bridge with equilibration of the two parameters [6]. The values of ΔR and ΔX must then be fed to a computing device which, incorrespondence with (8) and (9), will output the values of the specific impedance and the diameter.

In order to estimate the relationship between the quantities ΔR and ΔX, and also to show graphically the character of their variation, we construct the curves of the function $\Delta X = f(\Delta R)$, with the quantity ma as the parameter.* The graphs will be constructed on the basis of expressions (4) and (5) for ma < 10, and expressions (6) and (7) for ma > 10.

We shall operate with the relative values of the added impedances, $\Delta R / X_{ef_0}$ and $\Delta R / X_{ef_0}$ is the solenoid's reactive impedance without the bar (3).

The expressions for the relative magnitudes of the added impedances have the form:

for all values of ma

$$\frac{\Delta R}{X_{ef_0}} = \frac{2v}{ma} \frac{\operatorname{ber}(ma)\operatorname{ber}'(ma) + \operatorname{bei}(ma)\operatorname{bei}'(ma)}{\operatorname{ber}^2(ma) + \operatorname{bei}^2(ma)}, \tag{10}$$

$$\frac{\Delta X}{X_{ef_0}} = \frac{2v}{ma} \frac{\operatorname{ber}(ma)\operatorname{bei}'(ma) - \operatorname{bei}(ma)\operatorname{ber}'(ma)}{\operatorname{ber}^2(ma) + \operatorname{bei}^2(ma)} - v; \tag{11}$$

for ma > 10,

$$\frac{\Delta R}{X_{ef_0}} = \frac{\sqrt{2}\, v}{ma}, \tag{12}$$

$$\frac{\Delta X}{X_{ef_0}} = \frac{\sqrt{2}\, v}{ma} - v. \tag{13}$$

In these expressions, $v = a^2/b^2$ is the solenoid's space factor.

The functions $\Delta X \; X_{ef_0} = f(\Delta R \; X_{ef_0})$ for v = 1.0, 0.64 and 0.36 and given in Fig. 2.

The curves obtained are very important for the basis of the eddy current method of measuring. They are the starting points for the construction of the functions which will be given below, and they also show graphically in what way one may measure simultaneously, and independently of one another, two bar parameters: specific impedance and diameter.

* In [3], Förster gave an analogous diagram for the real and imaginary components of the voltage on the measuring coil.

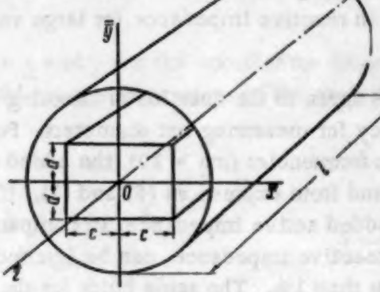


Fig. 5

Let us consider certain special features of the curves of Fig. 2 which render their construction easier.

1. The functions $\Delta X/X_{ef0} = f_1(\Delta R/X_{ef0})$, for ma = const and v = var, are straight lines from the origin of coordinates at the angle

$$\alpha = \operatorname{arc\,tg} \frac{C-1}{A}, \tag{14}$$

where

$$C = 2\frac{\operatorname{ber}(ma)\operatorname{bei}'(ma) - \operatorname{bei}(ma)\operatorname{ber}'(ma)}{ma\,[\operatorname{ber}^2(ma) + \operatorname{bei}^2(ma)]},$$

$$A = 2\frac{\operatorname{ber}(ma)\operatorname{ber}'(ma) + \operatorname{bei}(ma)\operatorname{bei}'(ma)}{ma\,[\operatorname{ber}^2(ma) + \operatorname{bei}^2(ma)]}.$$

2. Curves for different space factors can be obtained from one curve by multiplication by the space factor. Indeed, from the similar triangles OAD and OBE (Fig. 2) and on the basis of (10) and (11), it follows that

$$\frac{OA}{OB} = \frac{\Delta X_1}{\Delta X_2} = \frac{\Delta R_1}{\Delta R_2} = \frac{v_1}{v_2}, \tag{15}$$

where ΔX_1 and ΔR_1 are the added impedances corresponding to space factor v_1 and Bessel function argument ma, and ΔX_2 and ΔR_2 are the quantities which correspond to space factor v_2 and the same argument, ma.

It thus suffices to compute, using the Bessel functions, the curve $\Delta X/X_{ef0} = f(\Delta R/X_{ef0})$ for $v_1 = 1$; the curves for other space factors can be constructed on the basis of the relationships

$$\Delta X_2 = \Delta X_1 v_2, \tag{16}$$

$$R_2 = \Delta R_1 v_2. \tag{17}$$

On the diagram of $\Delta X/X_{ef0} = f(\Delta R/X_{ef0})$ (Fig. 2), each point unambiguously defines the specific impedance of the bar material and the bar's diameter. By using the diagrams of Fig. 2, with a sufficiently dense network of curves, one can obtain the values of γ and $\underline{a}$ by measuring the added impedance of the solenoid and knowing the solenoid characteristics and the frequency of the current.

It is important to note that changes in the specific conductivity and diameter of the bar shift the points characterizing the bar on the diagram of $\Delta X/X_{ef0} = f(\Delta R/X_{ef0})$ in different directions, separated by an angle of $\sim \pi/4$ (for ma $\leq$ 4). This fact might be used by employing two phase-sensitive elements, and graduating one instrument in units of conductivity and the other in units of length.

Based on expressions (10)-(13) and on Fig. 2, the curves of $\Delta R/X_{ef0} = f_1$ (ma) and $\Delta X/X_{ef0} = f_2$ (ma) were constructed on Figs. 3 and 4 respectively, showing, in another scale, the functions $\Delta R/X_{ef0} = \varphi_1(\gamma)$ and $\Delta X/X_{ef0} = \varphi_2(\gamma)$.

These curves permit a number of conclusions to be drawn.

1. To obtain maximum sensitivity, it is necessary to try to carry out the measurements with the largest possible space factor for the solenoid.

2. From the point of view of sensitivity in measuring specific impedance, the most useful values of ma are between 1.3 and 2.3.

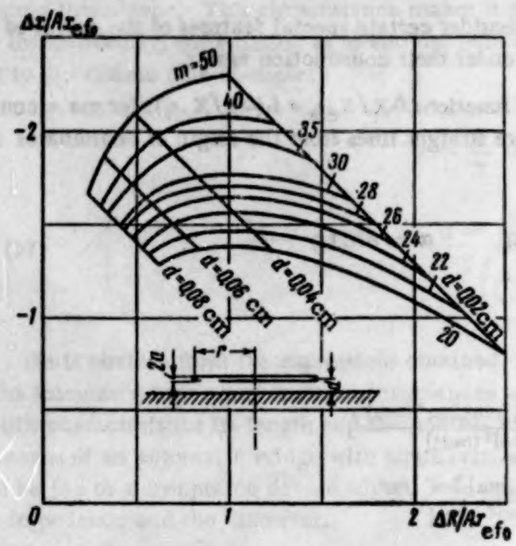


Fig. 6

3. For measuring diameter, it is convenient to use the changes in reactive impedance for large values of ma.

Let us turn again to the question of choosing the optimal frequency for measuring bar diameters. For sufficiently high frequencies (ma > 10), the added impedances are found from expressions (6) and (7). If ma > 140, the added active impedance, in comparison with the added reactive impedance, can be ignored with an error less than 1%. The same holds for the first term of expression (7).

Thus, we obtain the approximate formula for determining the radius of the bar:

$$a \approx \frac{1}{W}\sqrt{\frac{l}{\pi\mu_0\omega}\,|\Delta X|}. \tag{18}$$

For measurements of bar diameters, it is convenient to choose values of ma > 140 since, in this case, the magnitude of the specific impedance of the bar material does not affect the results of measuring the diameter. With this, the sensitivity, $d\Delta X$ da, is maximal.

III. Inspection of Prismatic Specimens and Blocks

We now find the expression for the desired impedances in the case when the specimen is prismatic in form. The necessity for such functions arose with the development of methods for inspecting the magnitude of the specific impedance by electrocarbon brushes [12].

With the introduction of a specimen, the total impedance of the solenoid is

$$\overline{Z} = R_{ef} + jX_{ef} = \frac{\dot{U}}{\dot{I}}, \tag{19}$$

where R_{ef} is the solenoid's active impedance, which is contributed to both by the solenoid's inherent active impedance at the given frequency, R_1, and by the heat (Joule effect) loss in the specimen, X_{ef} is the solenoid's active impedance, account being taken of the change in magnetic flux in the solenoid due to the presence of the specimen, $\dot{U}$ is the voltage supplied to the solenoid and $\dot{I}$ is the current flowing through the solenoid.

The voltage supplied to the solenoid is defined by the expression

$$\dot{U} = \dot{I}R_1 + \frac{d\dot{\psi}}{dt} = \dot{I}R_1 + j\omega\dot{\Phi}W, \tag{20}$$

where $\dot{\psi}$ is the solenoid's flux linkage, t is time, $\dot{\Phi}$ is the solenoid's magnetic flux, $\dot{\Phi} = \int_s \mu \dot{H} ds$, H is the magnetic field strength on area element ds of the solenoid's cross-section, s is the solenoid's cross-sectional area.

For a long solenoid and specimen, the magnetic field in the air gap between the solenoid turns and the specimen may be found from the formula

$$\dot{H}_0 = \frac{\dot{I}W}{l}. \tag{22}$$

The distribution of the magnetic field along a four-sided prismatic specimen placed in a long solenoid (Fig. 5) was found by G. A. Rasorenov [7]:

$$\dot{H}_m = \frac{\dot{I}W}{l}\left[\frac{\cos ky}{\cos kd} - \frac{2k^2}{d}\sum_{n=1}^{\infty}\frac{(-1)^n \cos \eta x}{\eta^2 \nu \cos \eta c}\cos \nu y\right], \tag{23}$$

where x and y are the coordinates the point of the prismatic specimen's cross section under consideration, 2c and 2d are the dimensions of the specimen's cross section,

$$k = \sqrt{j\mu_0\omega\gamma},$$

$$\nu = \frac{\pi}{2}\,\frac{1}{d}\,(2n-1) \quad (n = 1, 2, 3, \ldots).$$

On the basis of (19)-(23), one can find the expression for the solenoid's impedance, and can then find the expressions for the impedance added to the solenoid when a prismatically shaped specimen is introduced into it:

$$\Delta R = \frac{2\sqrt{2}\,\mu_0\omega c W^2}{ml}\left[\frac{\sin(\sqrt{2}\,md) - \operatorname{sh}(\sqrt{2}\,md)}{\cos(\sqrt{2}\,md) + \operatorname{ch}(\sqrt{2}\,md)}\right] - \frac{32\mu_0\omega d m^2 W^2}{\pi^2 l}\sum_{n=1}^{\infty}\frac{\sin\alpha\cos 3\varphi + \operatorname{sh}\beta\sin 3\varphi}{\lambda^3(2n-1)^2(\cos\alpha + \operatorname{ch}\beta)}, \tag{24}$$

$$\Delta X = -4\frac{\mu_0\omega c d W^2}{l} + \frac{2\sqrt{2}\,\mu_0\omega c W^2}{ml}\left[\frac{\sin(\sqrt{2}\,md) + \operatorname{sh}(\sqrt{2}\,md)}{\cos(\sqrt{2}\,md) + \operatorname{ch}(\sqrt{2}\,md)}\right] - 32\frac{\mu_0\omega m^2 d W^2}{\pi^2 l}\sum_{n=1}^{\infty}\frac{\operatorname{sh}\beta\cos 3\varphi - \sin\alpha\sin 3\varphi}{\lambda^3(2n-1)^2(\cos\alpha + \operatorname{ch}\beta)}. \tag{25}$$

Here,

$$\alpha = 2c\lambda\cos\varphi, \quad \beta = 2c\lambda\sin\varphi,$$

$$\lambda = \sqrt[4]{\frac{\pi^4(2n-1)^4}{16d^4} + m^4}, \quad \varphi = \frac{1}{2}\operatorname{arc\,tg}\left[-\frac{4m^2d^2}{\pi^2(2n-1)^2}\right].$$

The computation of expressions (24) and (25) does not entail any difficulty, since the sums entering into these expressions converge rapidly and, in practice, may be limited to the first three terms.

We now give the formulae relating the magnitudes of the impedances added to a circular (toroidal) coil with the specimen's parameters when the specimen is a block with a plane surface and the coil is above, and parallel to, this surface (Fig. 6). These formulae were obtained on the basis of the rules of inductance increase which were formulated by Wheeler [8]:

$$\Delta X = \frac{\mu_0 r\omega W^2}{d\sqrt{2\omega\mu_0\gamma}} - \mu_0 r\omega\left(\ln\frac{4r}{d} - 2\right)W^2, \tag{26}$$

$$\Delta R = \frac{\mu_0 r\omega W^2}{d\sqrt{2\omega\mu_0\gamma}}. \tag{27}$$

Here, r is the coil's radius, d is the distance from the mean height of the coil to the block. Formulae (26) and (27) are valid for the case when 2d $\ll$ r and for very high frequencies, when the depth of permeation is at least two times less that the thickness of the block.

If we transfer over to relative magnitudes (by dividing expressions (26) and (27) by the coil's inductive impedance when no specimen is present

$$X_{ef\cdot} = \mu_0 r\omega W'^2\left(\ln\frac{8r}{a} - 2\right) = \frac{1}{A}\mu_0 r\omega W^2, \tag{28}$$

where a is the radius of the solenoid's cross section) we can construct the curves

$$\frac{\Delta X}{AX_{ef_0}} = f\left(\frac{\Delta R}{AX_{ef_0}}\right)$$

for various values of d and $m = \sqrt{2}\,\mu_0\omega\gamma$ (Fig. 6). By going to relative quantities and introducing the quantity $A = [\ln(8r/d) - 2]^{-1}$, we make the curves of Fig. 6 universal, i.e., the curves become valid for coils of different dimensions.

It is clear from these curves that, for specimens with extended plane surfaces, eddy current measurements allows information to be obtained simultaneously with respect to two parameters, the specimen's conductivity and the distance of the conducting surface of the specimen from the coil transducer. This latter makes it possible to measure the depth of the nonconducting plating on metals, and also to construct very sensitive contactless displacement transducers.

IV. Sources of Error

In conclusion, we consider briefly the sources of the errors which can occur with eddy current methods of measuring.

First, the accuracy of measurement is determined by the accuracy of the instruments used to determine the magnitude of the added impedances ac bridges, various differential circuits, etc.

Second, inasmuch as the knowledge of the current frequency is necessary in determining the specific impedance and geometric dimensions of the specimen, the accuracy of the measurements will also be determined by the accuracy with which the generator supplying the measuring scheme maintains its frequency.

Third, there are errors due to deviations of the specimen from its normal position with respect to the coil transducer. For cylindrical specimens, this will be the effect of noncoaxiality of coil and specimen, and for plane specimens, the deviations from parallelism of the coil and specimen planes.

It is also necessary to take into account that we made a number of assumptions in the derivation of the formulae for the added impedances: we assumed that the solenoid and the cylindrical and prismatic specimens were very long; for the plane specimens, we assumed that the magnetic field was uniformly distributed in a surface layer of width $\epsilon/2$; we did not take into account the distributed capacitance in the coils, etc. All this leads to a definite error in the computations. To increase the accuracy of measurement, the instruments based on the use of eddy currents must be graduated according to the scale of the specimens.

SUMMARY

There is considered the noncontact method for measuring electric resistivity of nonmagnetic materials and geometric dimensions of massive samples of such materials. The method is based on use of eddy currents. Relationships are given which connect the introduced resistance values with the parameters of the samples when they have the cylindrical form, or the prismatic form, or the form of a flat surface plate.

LITERATURE CITED

[1] G. G. Yarmol'chuk, "Automation of processes in the production of electrocarbon goods," [in Russian]. Otchet IAT AN SSSR (1953).

[2] G. G. Yarmol'chuk, "Contactless methods of determining specific electrical impedance," [in Russian]. Automation and Remote Control (USSR) 19, 3 (1958).

[3] F. Förster, "Theoretical and experimental basis for nondestructive inspection with eddy currents," [in German]. Zeitschrift für Metallkunde, vol. 43, 4, 5, 6 (1952).

[4] R. Hochshild, Eddy Current Testing: A New Tool Makes Inspection Automatic Control Engineering, vol. 1, October, 1954.

[5] H. B. Dwight and M. M. Bagai, Calculation for Coreless Induction Furnaces. Trans. AIEE, vol. 54, 1935.

[6] V. Yu. Kneller, "AC bridges with automatic equalizing of two parameters," [in Russian]. Otchet IAT AN SSSR (1956).

[7] G. A. Razorenov, "On the distribution of fields and currents close to sharp angles between specimen boundaries," [in Russian]. Otchet Énergeticheskogo In-ta, AN SSSR (1941).

[8] H. A. Wheeler, Formulas for the Skin Effect. Proc. IRE, vol. 30, No. 9 (1942).

Received September 5, 1958

SIMULATION OF CHOKE-COUPLED MAGNETIC AMPLIFIERS

V. P. Glukhov and E. A. Yakubaitis

(Riga)

The similarity conditions of choke-coupled magnetic amplifiers are considered. Based on an analysis of the differential equations, similarity criteria are derived for amplifiers with ac windings connected in series, or in parallel.

The possibility is demonstrated of simulating amplifiers on one model with cross made of any electrotechnical steel. Results are given of the comparison of characteristics obtained from the model, and experimentally.

In the magnetic circuits of choke-coupled magnetic amplifiers there is a nonlinear and not well-defined relationship between induction and field strength. In connection with this, the integration of the differential equations describing the processes in such amplifiers is a difficult problem which becomes tractable only when significant assumptions are made.

The methods of solving the differential equations of the magnetic amplifier circuits reduce to methods of expressing, analytically or graphically, the induction of each of the amplifier's magnetic circuits as a function of its magnetic field strength. Such methods of expressing the magnetization curves, due to the assumptions used in making them, significantly lower the accuracy of the solutions.

Methods of designing amplifiers which are based on physical simulation [1-3] are free of the disadvantages mentioned.

For physical simulation, it is also necessary to work out mathematically the processes occurring in the system being investigated. However, the system equations can be written in a general form which is not applicable in an analytic solution.

The analytical (or graphical) solution of equations is now replaced by the experimental determination, on a model similar to the original being investigated, of the necessary functions.

1. A Magnetic Amplifier with Series-Connected ac Windings

If we ignore the leakage of the amplifier windings, the switching mode of the ideal rectifier in the feedback winding, the magnetic losses and the nonuniformity of induction distribution over the magnetic circuits' cross section, we may then write, for the ac and control circuits of an amplifier with series-connected ac windings (Fig. 1).

$$\sqrt{2}E_a \sin 2\pi ft = i_a R_a + L_L \frac{di_a}{dt} + W_a S\left(\frac{dB_A}{dt} + \frac{dB_B}{dt}\right)10^{-8} + + W_{fb} S\left(\frac{dB_A}{dt} - \frac{dB_B}{dt}\right)10^{-8}, \quad (1)$$

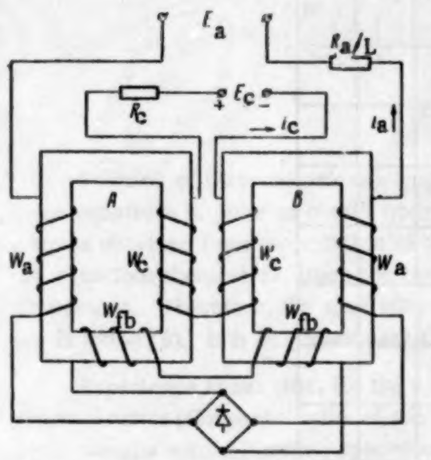


Fig. 1. Circuit of a magnetic amplifier with series-connected ac windings.

$$E_c = i_c R_c + W_c S\left(\frac{dB_A}{dt} - \frac{dB_B}{dt}\right)10^{-8}, \quad (2)$$

where E_a is the current value of voltage at the input to the ac circuit, i_a is the instantaneous value of current in that circuit, R_a is the active impedance of the ac circuit, L_L is the load inductance, W_a is the number of turns in the ac winding, E_c is the magnitude of the constant voltage applied to the input of the control circuit, i_c is the instantaneous value of current in that circuit, R_c is the total active impedance in the amplifier's control circuit, W_c is the number of turns in the control winding, S is the cross section of the magnetic circuit, B_A and B_B are the instantaneous values of induction in magnetic circuits A and B, $2\pi f = \omega$ is the angular frequency of the voltage at the input of the ac circuit, W_{fb} is the number of turns in the feedback winding and t is time.

After transformation, equation (1) may be written in the form

$$\sqrt{2}E_a \sin 2\pi ft = i_a R_a + L_L\frac{di_a}{dt} + W_a S\Big[(1 + k_{fb})\frac{dB_A}{dt} + + (1 - k_{fb})\frac{dB_B}{dt}\Big]10^{-8}, \quad (1a)$$

where $k_{fb} = W_{fb}/W_a$ is the feedback factor.

In each magnetic circuit, the induction is determined by the field strength induced by the ac circuit current, i_a, and the control circuit current, i_c:

$$H_A = \frac{W_a i_a + W_c i_c \pm W_{fb} i_a}{l}, \quad (3)$$

$$H_B = \frac{W_a i_a - W_c i_c \mp W_{fb} i_a}{l}, \quad (4)$$

where l is the average length of the magnetic circuit.

The dependence of the induction of the field strength of the magnetic circuit, in its general form, is given by the formula

$$B = \mu H, \quad (5)$$

where μ is the instantaneous value of the magnetic permeability of the material making up the magnetic circuit.

In its turn, the magnetic permeability [3] is a function of the magnetic circuit's field strength.

$$\mu = \mu_{bas}\varphi\left(\frac{H}{H_{bas}}\right), \quad (6)$$

where μ_{bas} and H_{bas} are the nominal base values of magnetic permeability and field strength respectively.

On the basis of equations (3)-(6), the expressions for the induction of the magnetic circuits can be written in the form

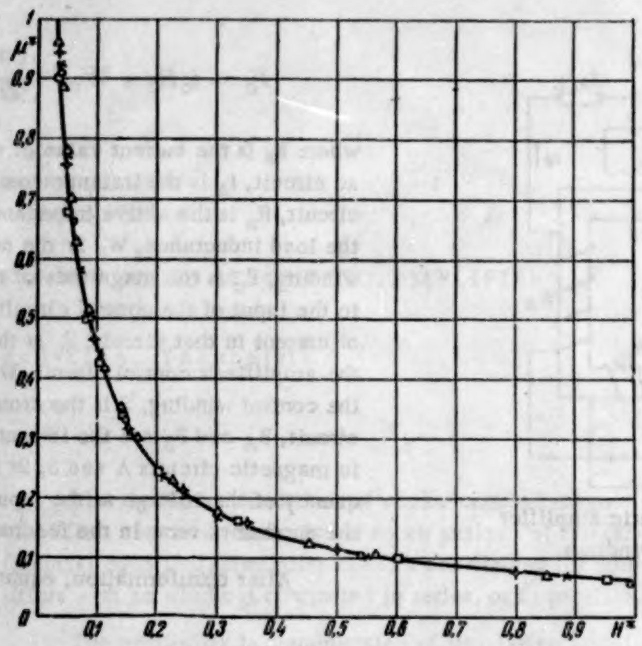


Fig. 2. Normal ac magnetization curve.

Symbol	Material	μ_{bas}, in gauss cm ampere	H_{bas}, in ampere cm
Circle	É330	60,000	4.92
Cross	50NP	54,000	5.00
Circle with cross	80NKhS	115,000	1.00
Triangle	N79M5	70,000	2.00
Square	É42	5,800	4.00

$$B_A = \mu_{bas} \frac{W_a i_a (1 \pm k_{fb}) + W_c i_c}{l} \varphi \left[\frac{W_a i_a (1 \pm k_{fb}) + W_c i_c}{H_{bas} l} \right], \tag{7}$$

$$B_B = \mu_{bas} \frac{W_a i_a (1 \pm k_{fb}) - W_c i_c}{l} \varphi \left[\frac{W_a i_a (1 \mp k_{fb}) - W_c i_c}{H_{bas} l} \right]. \tag{8}$$

By substituting expressions (7) and (8) in equations (1a) and (2), we may write:

$$\begin{aligned} \sqrt{2} E_a \sin 2\pi f t = i_a R_a + L_L \frac{di_a}{dt} + W_a S \mu_{bas} (1 + k_{fb}) \times \\ \times \frac{d}{dt} \left\{ \left[\frac{W_a i_a}{l} \left(1 \pm k_{fb} + \frac{W_c i_c}{W_a i_a} \right) \right] \varphi \left[\frac{W_a i_a}{H_{bas} l} \left(1 \pm k_{fb} + \frac{W_c i_c}{W_a i_a} \right) \right] + \right. \\ \left. + \frac{1 - k_{fb}}{1 + k_{fb}} \left[\frac{W_a i_a}{l} \left(1 \mp k_{fb} - \frac{W_c i_c}{W_a i_a} \right) \right] \varphi \left[\frac{W_n i_n}{H_{bas} l} \left(1 \mp k_{fb} - \frac{W_c i_c}{W_a i_a} \right) \right] \right\} 10^{-8}, \end{aligned} \tag{9}$$

$$E_c = i_c R_c + W_c S \mu_{bas} 10^{-8} \frac{d}{dt} \left\{ \frac{W_a i_a}{l} \left(1 \pm k_{fb} + \frac{W_c i_c}{W_a i_a} \right) \right] \times \tag{10}$$

$$\times\varphi\left[\frac{W_a i_a}{H_{bas} l}\left(1\pm k_{fb}+\frac{W_c i_c}{W_a i_a}\right)\right]-\left[\frac{W_a i_a}{l}\left(1\mp k_{fb}-\frac{W_c i_c}{W_a i_a}\right)\right]\times$$
$$\times\varphi\left[\frac{W_a i_a}{H_{bas}}\left(1\mp k_{fb}-\frac{W_c i_c}{W_a i_a}\right)\right]\bigg\}. \qquad (10)$$

Solution of these equations requires that function (6) be found. However, it is not necessary to integrate these equations in order to obtain from them the similarity conditions necessary for simulation. The similarity criteria obtained from the solution of the equations can not be different from the criteria obtained directly from the equations themselves, since the operation of solving an equation can not change the property of similar phenomena. Therefore, the similarity criteria may be found directly from equations (9) and (10) without finding function (6). It is only necessary that this function be the same in the model and in the original.

Experience shows that, for the corresponding choices of base quantities, μ_{bas} and H_{bas}, the magnetization characteristics (the mean value of the magnetic permeability as a function of the mean field strength) for magnetic circuits with differently spaced cores and realized by different ferromagnetic materials may, in relative units, be reduced to one curve, $\mu^* = \varphi(H^*)$. Thus, for example, Fig. 2 shows the magnetization characteristics for magnetic conductors made of ferromagnetic materials É330, 50NP, 80NKhS and N79M5 reduced to the magnetization curve of a toroidal magnetic circuit prepared of type É42 electrotechnical steel. Here are also given comparisons of the relative characteristics of the magnetic conductors, realized by one and the same material but having different core shapes (toroidal and pi-shaped) with different air gaps (δ = 0 and δ = 0.3 mm).

As the base magnitude of magnetic permeability we take the maximum value of permeability of the given ferromagnet. The base value of magnetic field strength is chosen by starting from the condition that the characteristic, $\mu^* = \varphi(H^*)$, of the various magnetic conductors coincide at the point of relative permeability taken as the minimum, μ^*_{min}. Therefore, the base value of field strength in each magnetic conductor is determined from the equation

$$H_{bas}=\left(\frac{H}{H^*}\right)_{\mu^*_{min}},$$

where H is the field strength (in absolute units) corresponding to the value of permeability μ^*_{min} and H^* is the field strength (in relative units), the choice of which is arbitrary but identical for all magnetic conductors.

As follows from Fig. 2 (here we have taken μ^* = 0.06 and H^* = 1), the reduction, by the method stated, of the magnetization characteristics of different ferromagnetic materials to one curve allows one to obtain a function in relative units which, by analogy with electrical machines, may be called the normal magnetization characteristic.

If losses in the magnetic circuits are ignored, what has been stated with respect to the function $\mu^* = \varphi(H^*)$ is expressed in the instantaneous values of (6). This allows one to have, in the original and in the model, not only different forms of the cores, but also different ferromagnetic materials.

Analysis of equations (9) and (10) [3] gives the following similarity criteria for the model and the original:

$$K_{I}=ft,\qquad K_{IV}=\frac{W_a^2 i_a S\mu_{bas}\cdot 10^{-8}}{E_a tl},\qquad K_{VII}=\frac{W_a i_a}{H_{bas} l},$$
$$K_{II}=\frac{i_a R_a}{E_a},\qquad K_{V}=K_{fb},\qquad K_{VIII}=\frac{i_c R_c}{E_c}, \qquad (11)$$
$$K_{III}=\frac{L_L i_a}{E_a t},\qquad K_{VI}=\frac{W_c i_c}{W_a i_a},\qquad K_{IX}=\frac{W_c W_a i_a S\mu_{bas} 10^{-8}}{E_c tl}$$

By carrying out some uncomplicated transformations, and by giving the instantaneous value of the control current, i_c, in the form of the algebraic sum of the constant component, E_c/R_c, and the instantaneous value of

the variable component of the even harmonics, i_{ca}, the similarity criteria in (11) can be written in a somewhat different form:

$$
\begin{aligned}
&K_1 = K_{\mathrm{I}} = ft, && K_5 = K_{\mathrm{V}} = k_{\mathrm{fb}},\\
&K_2 = K_{\mathrm{II}} = \frac{i_a R_a}{E_a}, && K_6 = \frac{K_{\mathrm{VII}}}{K_{\mathrm{II}}} = \frac{E_a W_a}{H_{\mathrm{bas}} R_a l},\\
&K_3 = \frac{K_{\mathrm{I}} K_{\mathrm{III}}}{K_{\mathrm{II}}} = f\frac{L_L}{R_a}, && K_7 = \frac{K_{\mathrm{VIII}}}{K_{\mathrm{II}} K_{\mathrm{VI}}} = \frac{E_c W_c R_a}{E_a W_a R_c},\\
&K_4 = \frac{K_{\mathrm{I}} K_{\mathrm{IV}}}{K_{\mathrm{II}}} = f\frac{W_a^2 S \mu_{\mathrm{bas}} \cdot 10^{-8}}{R_a l}, && K_8 = \frac{K_{\mathrm{IV}} K_{\mathrm{VIII}}}{K_{\mathrm{II}} K_{\mathrm{VI}} K_{\mathrm{IX}}} = \frac{R_c W_a^2}{R_a W_c^2},
\end{aligned}
\tag{12}
$$

$$
K_9 = \frac{i_{ca} W_c R_a}{E_a W_a}.
$$

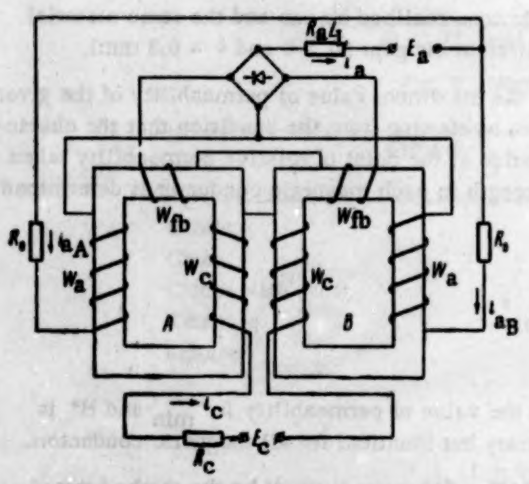


Fig. 3. Circuit of a magnetic amplifier with parallel-connected ac windings.

Consequently, in order that the model be similar to the original, it is necessary that criteria K_1-K_9 be identical for the model and the original. For investigations of transient responses, it is also necessary that the boundary conditions be similar.

Thus, we have nine similarity criteria and the two equations relating them (9) and (10). Therefore, two criteria are indeterminate (for the analysis of steady-state processes, these are K_2 and K_9). The remaining seven similarity criteria contain 14 parameters. Consequently, seven model parameters may be chosen arbitrarily, and will remain invariable in the model, independently of the parameters of the original. The remaining seven model parameters must vary in such fashion that, depending on the given parameters of the original and the selected seven parameters, the model will maintain all criteria invariant. Consequently, among these parameters must be those which would guarantee this invariance.

As shown by an analysis of the criteria in (12), as the invariant model parameters may be chosen S, l, W_a, W_c, f, H_{bas} and μ_{bas}. In this case, each defining criterion contains just one variable parameter (in the same order as the criteria): t, i_a, L_L, R_a, k_{fb}, E_a, E_c and R_c.

Thus, for correct simulation, the model and the original must provide equal values of the following quantities.

1. The product of frequency and time. Thus, the time scale chosen for the simulation depends on the ratio of frequencies in the model and original. Criterion K_1, as given in the literature [3], is called the criterion of homochronism, or homogeneity in time.

2. The ratio of the product of frequency and inductive load to the active impedance of the ac circuit.

3. The relationship between the design parameters, defined by criterion K_4.

4. The feedback factor, k_{fb}.

5. The ratio of the maximum possible field strength to the base field strength.

6. The degree of dc excitation of the amplifier.

7. **The ratio of the active impedances in the dc and ac circuits, reduced to one of these circuits. These ratios are usually called [4] coefficients of even harmonics suppression; the ratios inverse to these are called power gains [5].**

2. A Magnetic Amplifier with Parallel-Connected ac Windings

For a magnetic amplifier in which the ac windings are connected in parallel (Fig. 3), we may write, analogously to equations (1) and (2),

$$\sqrt{2}E_a \sin 2\pi ft = i_a R_L + L_L \frac{di_a}{dt} + i_{a_A} R_0 + W_a S \times \tag{13}$$

$$\times \left[(1 + k_{fb}) \frac{dB_A}{dt} - k_{fb} \frac{dB_B}{dt} \right] 10^{-8},$$

$$\sqrt{2}E_a \sin 2\pi ft = i_a R_L + L_L \frac{di_a}{dt} + i_{a_B} R_0 + W_a S \times \tag{14}$$

$$\times \left[k_{fb} \frac{dB_A}{dt} + (1 - k_{fb}) \frac{dB_B}{dt} \right] 10^{-8},$$

$$E_c = i_c R_c + W_c S \left(\frac{dB_A}{dt} - \frac{dB_B}{dt} \right) 10^{-8}, \tag{15}$$

where R_0 is the active impedance of the amplifier's ac winding, R_L is the active impedance of the load and the feedback winding.

It follows from equations (13) and (14) that

$$\sqrt{2}E_a \sin 2\pi ft = i_a R_a + L_L \frac{di'_a}{dt} + W_a S_{ha} \left[(1 + 2k_{fb}) \frac{dB_A}{dt} + \right. \tag{16}$$

$$\left. + (1 - 2k_{fb}) \frac{dB_B}{dt} \right] 10^{-8},$$

where

$$i_a = i_{a_A} + i_{a_B}, \tag{17}$$

$$S_{ha} = \frac{S}{2}, \quad R_a = R_L + \frac{R_0}{2}. \tag{18}$$

The field strengths of the magnetic circuits are defined by the expressions

$$H_A = \frac{W_a i_{a_A} - W_c i_c \pm W_{fb} (i_{a_A} - i_{a_B})}{l}, \tag{19}$$

$$H_B = \frac{W_a i_{a_B} - W_c i_c \mp W_{fb} (i_{a_A} + i_{a_B})}{l}. \tag{20}$$

By using equations (5) and (6), we may write the expressions for the magnetic induction of the magnetic circuits in the following form:

$$B_A = \mu_{bas} \frac{(1 \pm k_{fb}) W_a i_{a_A} \pm k_{fb} W_a i_{a_B} + W_c i_c}{l} \times$$

$$\times \varphi\left[\frac{(1 \pm k_{fb})\, W_a i_{a_A} \pm k_{fb} W_a i_{a_B} + W_c i_c}{H_{bas} l}\right],$$

$$B_B = \mu_{bas} \frac{(1 \mp k_{fb})\, W_a i_{a_A} \mp k_{fb} W_a i_{a_B} - W_c i_c}{l} \times$$

$$\times \varphi\left[\frac{(1 \mp k_{fb})\, W_a i_{a_A} \mp k_{fb} W_a i_{a_B} - W_c i_c}{H_{bas} l}\right].$$

The expressions just obtained may be written in a somewhat different form

$$B_A = \mu_{bas} \frac{W_a i_a (q \pm k_{fb}) + W_c i_c}{l}\, \varphi\left[\frac{W_a i_a (q \pm k_{fb}) + W_c i_c}{H_{bas} l}\right], \tag{21}$$

$$B_B = \mu_{bas} \frac{W_a i_a (q \mp k_{fb}) - W_c i_c}{l}\, \varphi\left[\frac{W_a i_a (q \mp k_{fb}) - W_c i_c}{H_{bas} l}\right], \tag{22}$$

where

$$q = \frac{i_{a_A}}{i}. \tag{23}$$

From a comparison of the differential equations of the magnetic amplifier circuits with series, and parallel, connections of the ac windings, it follows that equations (15) and (16) are analogous to equations (2) and (1a). The difference consists only in this, that with parallel connections, the cross section of the magnetic circuit and the active impedance of the ac winding both decrease, but the feedback factor is doubled. The expressions for the inductance of the magnetic circuits (7), (8) and (21), (22) differ by the presence in the latter of the factor q. Consequently, for similarity of the model and the original when the ac windings are connected in parallel, it is necessary that, in addition to condition (16), there be maintained the ratio of the current scales

$$\frac{m_{i_{a_A}}}{m_{i_a}} = 1.$$

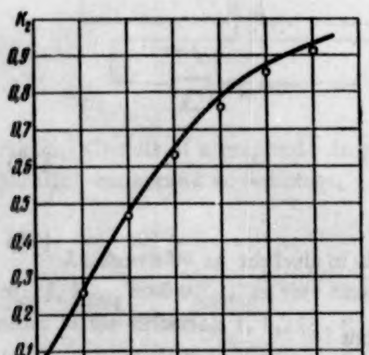


Fig. 4. Magnetic amplifier characteristics and the experimental points obtained from the model.

The current in one of the magnetic circuits of the amplifier with parallel-connected ac windings is a function of time, $i_{a_A} = f(t)$.

Since magnetic circuit B in no way differs from magnetic circuit A, but a periodic emf, E_a, is impressed on the input of the ac loop (Fig. 3), there will be the steady-state real function for the current i_{a_B}, $i_{a_B} = -f(t + \pi)$. In this case, the total current in the ac circuit, in accordance with (17) will be $i_a = f(t) - f(t + \pi)$.

Thus, as the scale of current i_{a_A} changes, there is a corresponding change in the scales of current i_{a_B} and the total current, i_a.

Consequently, relationship (24) holds for any parameters of the magnetic amplifier circuits, and the criteria in (11) determine the similarity conditions both for series and parallel connections of the ac windings.

The accuracy of the magnetic amplifier simulation is illustrated in Fig. 4. Here, in terms of the criteria, are shown the characteristics of a magnetic amplifier with toroidal cores of N79M5 Permalloy. The points on this figure give the same function, determined from the model, an amplifier with cores of É42 electrotechnical steel. Figure 4 shows that that simulation allows the characteristics of the original to be reproduced with sufficiently high accuracy on the model.

SUMMARY

1. In simulating a choke-coupled amplifier circuit, the simulation must extend both to the magnetic amplifier itself and to the measuring device and load on which the amplifier operates.

2. With the assumptions made in this paper, one and the same model with invariant parameters S, l, W_a, W_c, H_{bas} and μ_{bas}, with a constant frequency, f, can model an original with arbitrary parameters for a variable frequency in the static modes of operation.

3. Similarity of choke-coupled magnetic amplifier circuits is determined by the nine criteria of (12).

4. With no feedback present, it is possible to simulate, on one and the same model, both parallel and series connections of the ac windings in the amplifier.

LITERATURE CITED

[1] M. V. Kirpichev, Similarity Theory. [in Russian]. Published by the AN SSSR (1953).

[2] L. S. Eigenson, Simulation [in Russian]. Sovetskaya Nauka (1951).

[3] V. A. Venikov. The Application of Similarity Theory and Physical Simulation to Electric Engineering [in Russian]. Gosénergoizdat (1949).

[4] G. F. Storm. Magnetic Amplifiers. [Russian translation] IL (1957).

[5] M. A. Rozenblat. Magnetic Amplifiers. [in Russian]. "Sovetskoe Radio" (1956).

Received February 24, 1958

THE EFFECT OF NOISE ON SYNCHRONOUS FILTER-GENERATOR OPERATION

A. M. Luchuk

(Kiev)

An analysis is made of the operation of a synchronous filter generator with sinusoidal noise at its input, and the results are given of an experimental investigation of the noise stability of telecontrol devices with synchronous filter generators. A qualitative and quantitative estimate of the effect of noise on synchronous filter operation is given, permitting correct calculation of its parameters.

In work [1] there is given a description of, and the theory of operation for, a selective filter for remote control frequency devices with narrow pass bands, called by the author a "synchronous filter generator" (SFG). In the operation of such a filter there will unavoidably be, at its input, in addition to the signal voltage, of frequency f_s, coinciding with the filter-generator frequency, f_g, an interfering signal (noise), U_n, of frequency f_n, different from frequency f_g or a multiple of it.

We give below an analysis of SFG operation with sinusoidal noise at its input. In the simplest case, such noise is a signal whose frequency differs from the filter-generator frequency when several signals are sent simultaneously. In addition, the results are given of the experimental investigation of noise stability of a telecontrol device with synchronous filters acted upon by random noise.

1. Analysis of the Effect of Sinusoidal Noise on SFG Operation

For calculations of the circuit for discriminating the synchronous filter signals (Fig. 1 in [1]), it is convenient to use the equivalent circuit, shown in Fig. 1. Ordinarily, the circuit parameters are such that no essential error is introduced into the calculations by idealizing the diode characteristics and by ignoring the internal impedances of the generator and the signal source.

At the circuit's input let there be, in addition to the signal voltage, a noise voltage of frequency f_n, different from the generator frequency f_g:

$$u_n = \sqrt{2}U_n \sin \omega_n t.$$

In this case, the instantaneous values of the voltages on loops I and II of the circuit will have the forms:

$$u_1 = \sqrt{2}U_s\,[n \sin(\omega_g t + \theta + \gamma) + \sin(\omega_g t + \gamma)] + \sqrt{2}U_n \sin \omega_g t,$$
$$u_2 = \sqrt{2}U_s\,[n \sin(\omega_g t + \theta + \gamma) - \sin(\omega_g t + \gamma)] - \sqrt{2}U_n \sin \omega_g t.$$

Here, $n = U_g/U_s$, ω_g and ω_n are the angular frequencies of generator and noise signals respectively,

θ is the initial phase shift of the voltages U_g and U_s and γ is the initial phase shift of the voltages U_g and U_n.

We introduce the notation

$$K_1 = \sqrt{1 + n^2 + 2n\cos\theta}, \quad K_2 = \sqrt{1 + n^2 - 2n\cos\theta},$$

$$m = U_n / U_s = m, \quad \omega_d = \omega_n - \omega_g = \omega_g,$$

$$\psi_1 = \operatorname{arc\,tg} \frac{n \sin\theta}{n\cos\theta + 1} + \gamma, \quad \psi_2 = \operatorname{arc\,tg} \frac{n\sin\theta}{n\cos\theta - 1}.$$

Then, after transformation, we obtain the following expressions for u_1 and u_2:

$$u_1 = \sqrt{2}\, U_1 \sin(\omega_g t + \beta_1), \quad u_2 = \sqrt{2} U_2 \sin(\omega_g t + \beta_2),$$

where

$$U_1 = U_s \sqrt{K_1^2 + m^2}\left[1 + \frac{2K_1 m}{K_1^2 + m_2}\cos(\omega_d t - \psi_1)\right]^{1/2},$$

$$U_2 = U_s \sqrt{K_2^2 + m^2}\left[1 + \frac{2K_2 m}{K_2^2 + m^2}\cos(\omega_d t - \psi_2)\right]^{1/2},$$

$$\beta_1 = \operatorname{arc\,tg} \frac{K_1 \sin\psi_1 + m\sin\omega_d t}{K_1\cos\psi_1 + m\cos\omega_d t}, \quad \beta_2 = \operatorname{arc\,tg} \frac{K_2\sin\psi_2 + m\sin\omega_d t}{K_2\cos\psi_2 + m\cos\omega_d t}.$$

On the impedances of the circuit, R_1 and R_2, will be discriminated the rectified components, multiples of ω_d and ω_g. The filter's output relay reacts to the difference of the constant rectified components, U_{01} and U_{02}. Let us determine these components.

For any values of K and m, the condition $2Km/(K^2 + m) < 1$ holds (ordinarily, $K > 2$). Therefore, an expression of the expression

$$\left[1 + \frac{2K_1 m}{K_1^2 + m^2}\cos(\omega_d t - \psi_1)\right]^{\frac{1}{2}} \quad \text{and} \quad \left[1 + \frac{2K_2 m}{K_2^2 + m^2}\cos(\omega_d t - \psi_2)\right]^{\frac{1}{2}}$$

by Newton's binomial formula gives a convergent series from which we obtain the formulae for U_{01} and U_{02}:

$$U_{01} = \frac{\sqrt{2}}{\pi} U_s \sqrt{K_1^2 + m^2}\left[1 - \frac{1}{4}\left(\frac{K_1 m}{K_1^2 + m^2}\right)^2 - \frac{15}{64}\left(\frac{K_1 m}{K_1^2 + m^2}\right)^4 - \dots\right],$$

$$U_{02} = \frac{\sqrt{2}}{\pi} U_s \sqrt{K_2^2 + m^2}\left[1 - \frac{1}{4}\left(\frac{K_2 m}{K_2^2 + m^2}\right)^2 - \frac{15}{64}\left(\frac{K_2 m}{K_2^2 + m^2}\right)^4 - \dots\right].$$

From these last expressions we determine the voltage at the filter output: $U_0 = U_{01} - U_{02}$.

Figure 2 gives the curves of the filter output voltage in relative units, U_0/U_s, as a function of m for various values of n and for $\theta = 0$, constructed from the equations just obtained. It follows from these curves that, for a given value of signal voltage, the appearance of noise at the filter input leads to a decrease in the voltage at its output. However, with increasing n, the lowering of the output voltage with noise at the filter

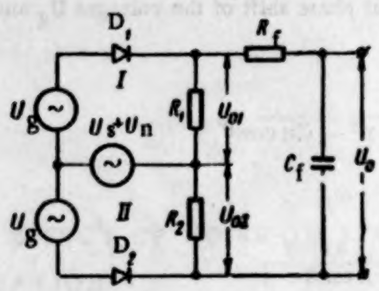


Fig. 1

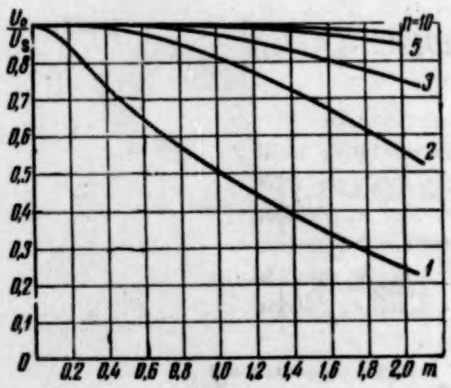


Fig. 2

input decreases, and even for n equal to 4 or 5, this lowering can be ignored in practice. Such a value of n should also be chosen for reliable filter operation in remote control devices.

The results obtained do not give a convenient method for determining the voltage at the circuit's output when the input is acted upon by a signal whose frequency is a multiple of the frequency of the local generator. Below, we give an analysis of circuit operation for this case.

Let the signal frequency be an even multiple of the generator frequency: $f_s = 2if_g$, where i is an arbitrary integer.

For $n > 1$ and $\theta = 0$, for the constant rectified components on impedances R_1 and R_2 we find

$$U_{01} = \frac{\sqrt{2}}{T} U_s \int_0^{T/2} (n \sin \omega_g t + \sin 2i\omega_g t)\, dt = \frac{\sqrt{2}}{\pi} U_s n,$$

$$U_{02} = \frac{\sqrt{2}}{T} U_s \int_0^{T/2} (n \sin \omega_g t - \sin 2i\omega_g t)\, dt = \frac{\sqrt{2}}{\pi} U_s n.$$

For the constant component of the output voltage we get

$$U_0 = U_{01} - U_{02} = 0.$$

Thus, the synchronous filter does not react to signal frequencies which are even multiples of the generator frequency. The analogous conclusion is easily obtained for the case when the filter input is composed of even subharmonics.

In a similar way we determine the effect on SFG operation of odd signal harmonics. As in the previous case, we have, for large n and $\theta = 0$,

$$U_{01} = \frac{\sqrt{2}U_s}{T} \int_0^{T/2} [n \sin \omega_g t + \sin (2i - 1)\, \omega_g t]\, dt,$$

$$U_{02} = \frac{\sqrt{2}U_s}{T} \int [n \sin \omega_g t - \sin (2i - 1)\, \omega_g t]\, dt,$$

which gives

$$U_{01} = \frac{\sqrt{2}}{\pi} U_s \left(n + \frac{1}{2i - 1}\right), \quad U_{02} = \frac{\sqrt{2}}{\pi} U_s \left(n - \frac{1}{2i - 1}\right),$$

$$U_0 = U_{01} - U_{02} = \frac{2\sqrt{2}}{\pi} U_s \frac{1}{2i - 1}.$$

Consequently, with odd signal harmonics at its input, the synchronous filter has an output voltage whose magnitude is inversely proportional to the multiplicity of the noise frequency. For $i = 1$ ($f_g = f_s$), the expression obtained coincides with the expression given in [1] for the maximum value of filter output voltage when its input signal coincides in frequency with the voltage of the local generator.

The conclusions obtained should be kept in mind in selecting the frequency range for a remote control device with synchronous filter generators. The effect of the third signal harmonic is the strongest, so it is

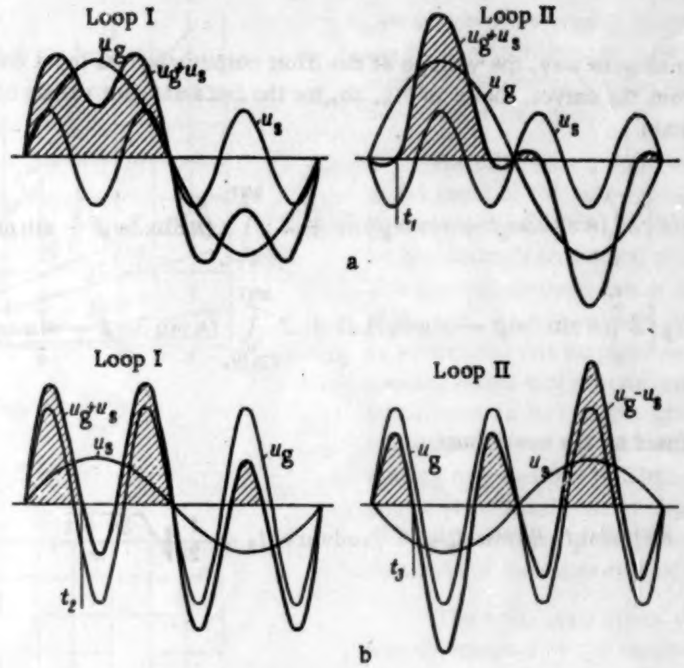


Fig. 3

advantageous to consider in more detail its influence on the filter's operation. The curves of the instantaneous values of the voltages in loops I and II of the circuit, in this case with small n, are shown in Fig. 3a.

For the constant rectified components we have

$$U_{01} = \frac{\sqrt{2}}{T} U_s \int_0^{T/2} (n \sin \omega_g t + \sin 3\omega_g t)\, dt,$$

$$U_{02} = \frac{\sqrt{2}}{T} U_s \left[\int_{t_1}^{T/4} 2\,(n \sin \omega_g t - \sin 3\omega_g t)\, dt + \int_{T/2}^{T/2+t_1} 2\,(n \sin \omega_g t - \sin 3\omega_g t)\, dt \right].$$

The limit of integration, t_1, is found from the condition

$$n \sin \omega_g t_1 - \sin 3\omega_g t_1 = 0,$$

which gives

$$\sin \omega_g t_1 = \frac{\sqrt{3-n}}{2}.$$

From this last equality it follows that the expressions given above for U_{01} and U_{02} are valid for $n \leq 3$. For $n > 3$, the expressions for U_{01} and U_{02} corresponding to the general case of odd signal harmonics remain in force.

For the given case, the constant component of the voltage at the filter output is defined by the expression

$$U_0 = U_{01} - U_{02} = \frac{2\sqrt{2}}{\pi} U_s \left[n - \frac{1}{3}(n+1)^{3/2} \right]$$

Analysis of this equation shows that, with a decrease of n from 3 to 1, the effect of the third harmonic on filter operation decreases.

We can find, in an analogous way, the voltage at the filter output when its input contains the third subharmonic of the signal. From the curves, shown in Fig. 3b, for the instantaneous values of the voltages in loops I and II in this case, we obtain

$$U_{01} = \frac{\sqrt{2}}{T} U_s \left[2 \int_0^{t_2} (n \sin 3\omega_s t + \sin \omega_s t)\, dt + 2 \int_{T/2+t_2}^{3/4T} (n \sin 3\omega_s t + \sin \omega_s t)\, dt \right],$$

$$U_{02} = \frac{\sqrt{2}}{T} U_s \left[2 \int_0^{t_3} (n \sin 3\omega_s t - \sin \omega_s t)\, dt + 2 \int_{T/2+t_3}^{3/4T} (n \sin 3\omega_s t - \sin \omega_s t)\, dt \right].$$

The value of t is defined by the conditions

$$n \sin 3\omega_s t_2 + \sin \omega_s t_2 = 0 \text{ and} \sin \omega_s t_2 = \frac{1}{2} \sqrt{\frac{3n+1}{n}}.$$

Analogously,

$$n \sin 3\omega_s t_3 - \sin \omega_s t_3 = 0 \text{ and} \sin \omega_s t_3 = \frac{1}{2} \sqrt{\frac{3n-1}{n}}.$$

Whence

$$U_0 = \frac{2\sqrt{2}}{\pi} U_s \left[1 - \frac{n+1}{3} \sqrt{\frac{n+1}{n}} + \frac{n-1}{3} \sqrt{\frac{n-1}{n}} \right].$$

The functions $U_0/U_{0\,max} = f(n)$, constructed from the above equations for action on the synchronous filter of the third harmonic and subharmonic signals, are given on Fig. 4. Here, $U_{0\,max}$ is the voltage at the filter output in the absence of noise.

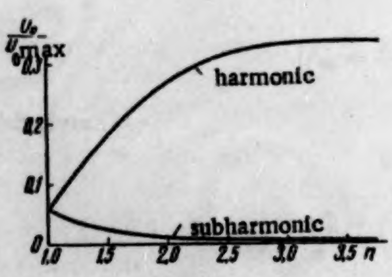


Fig. 4

From the functions given it follows that the signal subharmonics have no essential effect on filter operation and, in practice, need not be taken into account.

Similar calculations may also be carried out for the case when higher harmonics and subharmonics act on the filter. However, as follows from the expression for the general case of odd harmonic action, their effect on filter operation is insignificant. Therefore, such calculations have no practical value, the more so as there is always the possibility, in actual remote control systems, using simple frequency correction devices which allow the amplitude of the signal harmonics and subharmonics to be lowered to the degree desired.

2. Experimental Noise Stability Determination for a Remote Control Device with SFG

With frequency methods of coding, noise can not essentially vary the signal form, but can give false signals. From the point of view of reliability of remote control device operation, the most dangerous is the false operation of a filter relay due to noise.

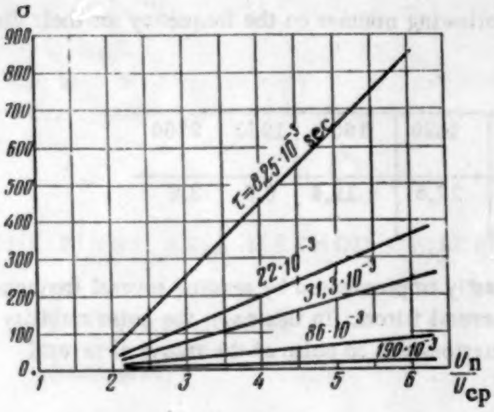


Fig. 5

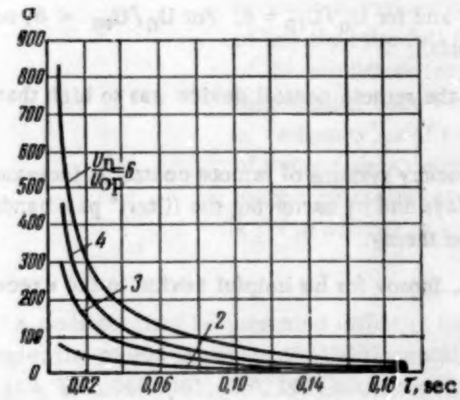


Fig. 6

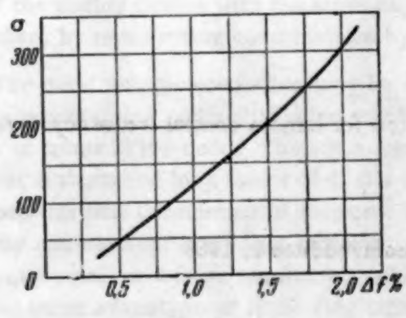


Fig. 7

For the purpose of verifying the reliability of operation of a remote-control frequency device with synchronous filter generators, an experimental investigation was made of the device's noise stability when random noise appeared at its input.

For this, from a noise source there was introduced at the input of the half-complete remote-control device a noise voltage which, after preliminary amplification by the device's amplifier, was fed to six SFG in parallel. The spectral composition of the noise had a random character. As a noise source we used a program, specially written for this purpose, on a magnetic sound recorder: speech, music and atmospheric noise, approximating white noise in its spectral characteristics. The frequency characteristics of the recorder were uniform in the frequency range of 300 to 3000 cycles. The device's amplifier provided distortion-free amplification for a four-fold to five-fold excess of the noise over the operation threshold of the filter relays.

The noise stability of the remote control device was determined by the number of operations, σ, of the filter's output relays by the noise.

The following relationships were determined:

1. The dependence of the number of operations of a filter's output relay on the voltage level of the noise, $\sigma = F(U_n \; U_{op})$, where U_{op} is the effective value of signal voltage at the device's input necessary to operate the filter relay when the signal frequency coincides exactly with the frequency of the local generator.

With a signal voltage of $U_s = 2U_{op}$, the operation band of the filter relay comprised, in this case, 1-6 of the frequency of the local generator. Figure 5 shows the dependencies obtained for various values of the filter's time constant, $\tau = R_f C_f$ (Fig. 1).

2. The number of operations of an output relay as a function of τ for the same relay operation band (Fig. 6).

3. The function $\sigma = F(\Delta f)$, where Δf is the filter's relay operation band for a signal voltage twice the voltage for filter relay operation (Fig. 7).

In determining the functions listed, one and the same program was applied to the device's input for different values of $U_n \; U_{op}$, τ and Δf. The length of one cycle was 32 minutes.

It follows, from the dependencies obtained, that for a given noise level, the noise stability of a remote control device with synchronous filter generators can be significantly increased by increasing the time constant τ and by lowering the filters' pass band down to a minimum value determined by stability considerations for the generator's frequency.

The distribution of the number of operations over the individual filters was determined by the spectral content of the noise. For the given program, the number of operations of the individual filter relays, as

percentages of the total number of operations, depended in the following manner on the frequency for their distribution:

Frequency in cycle	820	1280	1420	1600	1970	2760
Number of operations, %	28.5	29.8	17.5	12.4	8.6	3.2

In remote control devices, the control operations are ordinarily implemented by sending several frequencies, which requires the simultaneous operation of the relays of several filters. In this case, the noise stability of the device was determined by the number of simultaneous operations due to noise of the relays of several filters.

It was established that, of 2810 cases of relay operation by noise, only seven cases occurred of simultaneous operation of two relays.

Simultaneous operation of three relays did not occur. It should be mentioned that all cases of two-relay operation were observed only for small τ (less than 0.031 seconds) and for $U_n/U_{0p} = 6$. For $U_n/U_{0p} < 5$, no cases of double operation were observed, even for $\tau = 0.008$ seconds.

Consequently, even for $\tau = 0.031$ seconds, noise stability of the remote control device was so high that normal operation of the device was not disturbed.

We may thus draw the conclusion that noise stability of frequency systems of remote control is increased by limiting the noise level, by increasing the lag of the filters' relays and by narrowing the filters' pass bands. This is in agreement with the statements of general communication theory.

In conclusion, the author wishes to express his thanks to V. L. Inosov for his helpful advice in the execution of the present work.

SUMMARY

Operation of a synchronous filter-generator with an input sinusoidal noise in analyzed. Results of the experimental treatment of noise stability of the telecontrol device with synchronous filter generators are described. Qualitative and quantitative estimating of noise influence on the synchronous filter operation is suggested that permits to correctly calculate the filter parameters.

LITERATURE CITED

[1] V. L. Inosov and A. M. Luchuk, "Synchronous filter generators for remote control frequency devices," [In Russian]. Automation and Remote Control (USSR) 17, 10 (1956).

Received June 9, 1958

CODE RINGS AS A METHOD OF REPRESENTING CODE SETS

A. N. Radchenko
(Leningrad)

A method is considered for the shortening of the writing of sets of code combinations by excluding from the constituents of various codes the repeated combinations of terms of less than the full (maximum) length. An analysis is made of the conditions for the existence of such shortened forms of representing code sets, called code rings, and the existence of various types of code rings is proved. The shortened form of writing code combinations may be used for decreasing the number of circuit elements or for increasing the capacity of a given number of elements.

A code set may be presented either in the form of a table or in the shorter form of a code ring [1]. For example, the closed sequence 01110001 contains all the possible three-place binary code combinations: 011, 111, 110, 100, 000, 001, 010, 101, each combination being represented once and only once. The number of elements necessary to write the code set by this method is sharply decreased. Thus, to write all the five-place binary codes in the form of a table requires 160 letters while only 32 letters are needed for the code-ring representation. This permits a five-fold decrease in the number of physical elements in the corresponding portions of the coding device with the same capacity being retained. If changing of the coding device is not necessary then, by representing combinations by code rings, one can increase the capacity proportionately.

The code ring representation may be used for any other code set, formed from a different number of symbols, $\underline{k}$, or containing codes with other numbers of terms, $\underline{n}$. In all cases, the shortening obtained equals the number of terms in the code. Thus, if a four-place code is used, the number of elements necessary to write the entire set is shortened by a factor of 4, if a ten-place code is used, a ten-fold decrease is obtained, etc. It follows from this that the number of elements in a code ring always equals the number of codes to be represented, i.e., only one element is used for representing any of these combinations in a ring, independently of whether the code uses a small or a large number of places. As a result, the greater the number of terms in the code to be used, the more advantageous is the ring representation of the code set.

Below, we shall consider the conditions for the existence of code rings for sets of number-positional codes for any $\underline{k}$ or $\underline{n}$. First, however, we shall define a number of concepts.

A collection of codes, formed in accordance with a stipulated rule, is a finite set for bounded $\underline{k}$ and $\underline{n}$, which we agree to call a code set. In the case of a number-positional code, the full set contains $m = k^n$ codes; a code set formed by a quality-shift rule (i.e., no two adjacent elements can be identical) consists of $m = k(k-1)^{n-1}$ codes; a set formed by taking $\underline{k}$ symbols (k-1) at a time consists of $m = {}_kC_{k-1} = k!$ codes, etc.

We agree to give the name "code ring" to a closed sequence of symbols so constructed that the number of elements in it equals the number of nonrepeating n-term combinations which are formed by the segments of the given sequence (including also the overlapping segments formed in closing the sequence).

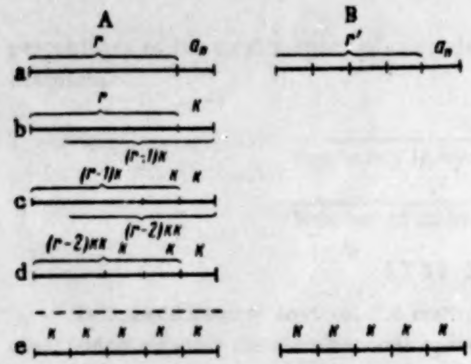


Fig. 1. For the proof of the impossibility of decomposing a complete number-positional set into nonintersecting parts.

To elucidate the conditions of existence of code rings, we use the property of codes which, in correspondence with algebraic terminology, we define as "intersection." Two codes are called intersecting if the combination of n-1 elements at the end of one code is identical to the combination of n-1 elements at the beginning of the other code, for example, the codes ababb and babbb intersect in the combination babb.

If the combinations of the set are laid out in a sequence of their intersections, then the case is possible when the extreme code combinations also intersect. If we thus order the code arrangements, we can obtain a ring. In correspondence with the definition of code ring already given, we shall call a closed sequence of symbols, established from the ordered sequence of code combinations, a code ring if the intersecting portions of each pair of adjacent codes enter into the sequence once and only once. From this comes the first necessary condition: for the existence of a ring representing the combinations of a given code set, it is necessary that, in the given code set, the right and left (n-1)-place combinations enter the same number of times.

It is easy to show that this condition holds identically for a complete set of number-positional codes. To show this, we present one such code in the following form

$$\left\{\rlap{\overbrace{\phantom{a_1a_2,\ldots,a_{n-1}}}^{n-1\text{ places}}}a_1\underbrace{a_2,\ldots,a_{n-1}a_n}_{n-1\text{ places}}\right\}$$

where $a_s = a, b, c, \ldots, k$ $(s = 1, \ldots, n)$.

By writing the individual incomplete combinations, corresponding to the right and left parts of the individual codes:

$$\{a_1 a_2, \ldots, a_{n-1}\} = \{a_2, \ldots, a_{n-1}, a_n\},$$

we convince ourselves of their identity, which is determined both by the identity of the number of terms and by the generality of the law of the independent combinations of k values of each of the terms (places). This follows identically from the fact that each fixed group of n-1 elements occurs on the right k times, inasmuch as to obtain all possible codes with the given group on the right one can write at its left only k different symbols. Since, in accordance with the rule for constructing a complete code set, it is indifferent whether a lacking term is written on the right or on the left, an equal number of codes (i.e., k) will have the given fixed group on the left. Thus, in a complete number-positional code set, it is possible to form k code pairs which intersect in a fixed group of elements, where any of the intersecting codes must enter into only one pair, if it is understood that a self-intersecting code* forms an independent pair.

The fact that there exists for each code another code which intersects it is not sufficient for the construction of a ring, since it is always possible to assume that a given set is decomposed into independent portions between which there are no intersections. Then, the given code set can not be represented by one ring, but only by a system of rings. From this derives the second necessary condition: for there to exist one ring representing all the

---

* By a self-intersecting code is meant a code of the type $\rlap{\overbrace{}^{n-1\text{ places}}}a\,\underbrace{a\,a\,a}_{n-1\text{ places}}$, the right incomplete portion of which is identical with the left portion.

code combinations of a given code set, it is necessary that it be impossible so to decompose the given set into two parts that one of the parts would not have at least one code which intersects a code contained in the other part.

We now prove that a complete code set of number-positional codes is impossible to decompose into two parts in such fashion that one of the parts would not contain at least one code which intersects a code in the other part.

To show this, we assume the converse, i.e., that the decomposition mentioned can be carried out, so that we obtain two code groups, A and B, between which there are no intersecting codes. We select from group A any of its codes, and denote by $\underline{r}$ its (n-1)-place left-most combination, and by a_n the last term (Fig. 1a). Then in accordance with the assumption made, group A must contain all codes whose right and left (n-1)-place combinations are identical with combination $\underline{r}$. This follows from the fact that the combination $\underline{r}$ cannot be found in group B either on the left or on the right, since otherwise these codes would intersect either with the initially considered code of group A or with those of the group A codes which intersect with the original one. If all the codes in which the left or right (n-1)-place combination is identical to $\underline{r}$ are in group A, then among these codes must be found such a code whose left combination is $\underline{r}$ and whose remaining term is $\underline{k}$ (Fig. 1b).

If, on Fig. 1b, we isolate the (n-1) elements at the right of the code and designate them by (r-1)k, we are led, analogously to the reasoning of the last paragraph, to the conclusion that in group A there must be found all the codes in which the (n-1)-place combinations are identical with (r-1)k. Among these must be the one in which the combination (r-1)k is on the left and whose right extreme is $\underline{k}$ (Fig. 1c). Thus, in group A there must be a code terminating with the two terms kk. Analogously, we arrive at the conclusion that the rightmost (n-1)-place combination must occur at the left of codes in group A. This, in its turn, determines the existence in group A of a code which terminates in three k's, etc. By continuing this reasoning we come unavoidably to the conclusion that group A contains a code consisting entirely of k's (Fig. 1e).

An analogous line of reasoning can be followed, starting with any code, including any code from group B. Thus will be proved the existence in group B of a code consisting exclusively of k's. The identical resulting codes, obtained from groups A and B, intersect each other, which contradicts both the initial assumption and the condition of nonrepetition of codes within a given set. Consequently, it is impossible to decompose the complete set of number-positional codes into two parts such that one of them does not contain at least one code which intersects a code lying in the other part.

The proof of the existence of code rings leads to the verification of the presence of the first and second conditions in the code set. Such an analysis was carried out above for the complete set of number-positional codes. By the same analytical path, one may show that other complete code sets can also be represented in the form of code rings. In particular, code rings can always be constructed for quality-shift codes, and for codes formed from the set of combinations (in the sense of "permutations and combinations"). The same can be said of permutations if, from the corresponding combinations, the last, completely redundant term is removed. In [1] were given examples of quality-shift and number-positional code rings, and also an example of a ring representation of the complete set of combinations.

Such rings, where all possible code combinations are represented just once within the cycle limits, we agree to call type A code rings. Type A code rings for number-positional codes may always be formed. However, the possibility is not excluded of forming code rings which only include parts of the codes. Cases do arise in which a complete code set can be decomposed into parts, each of which can be represented by an independent code ring. As in the general case, each code which enters into a given portion occurs in such (partial) code rings only once. The total number of elements necessary to represent a code set in the form of a system of partial code rings is the same as with the use of a type A code ring. Analogously, the number of elements in the system of partial code rings equals the number of codes and does not depend on the number of terms (places) in the code.

In correspondence with the second condition, there must be contained, in any partial code ring of a number-positional code, at least one code which intersects with a code of another partial code ring. By the use of this property, the partial code rings can be fused into one type A code ring. An effective method of constructing type A code rings is developed on this basis. A type A code ring possesses a number of specific properties. For example, by counting off the ring elements in reverse order, we obtain a new code ring of the same type, which differs from the given one in the order in which its constituent combinations follow one another. This property

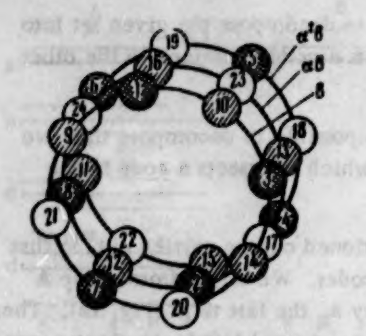


Fig. 2. A grouped code ring of a quality shift code. $k = 3$, $n = 4$, $m = k(k-1)^{n-1} = 24$. 1) abab, 2) abac, 3) abca, 4) abcb, 5) acab, 6) acac, 7) acba, 8) acbc, 9) baba, 10) babc, 11) baca, 12) bacb, 13) bcab, 14) bcac, 15) bcba, 16) bcbc, 17) caba, 18) cabc, 19) caca, 20) cacb, 21) cbab, 22) cbac, 23) cbca, 24) cbcb.

is easily proven by reading any complete code table from right to left. It may be established by this that all combinations of the given code set are also duplicated in such a description. Consequently, when the direction of reading is reversed in a ring, all codes of the given code group must again be observed, and only once each.

The second property consists of the fact that transposition of any two symbols of a code ring also transforms it into a ring of the same type, differing only in the order in which the codes in the code ring follow one another. This property may be proven if the transposition of the two symbols is carried out for all the codes which contain them. With this, none of the codes disappears from the table. This latter is explained by the circumstance that in any complete group of codes it is possible to isolate pairs of codes in such wise that, upon transposition, each code in a pair goes over to its conjugate.

By a repeated application of the second property, one can also observe the third property: any permutation of the symbols in a code ring does not change the type of ring, but changes only the arrangement of the codes in the ring.

By using the properties just enumerated, one can, by means of the corresponding operations, transform an initial code ring to a form most useful for practical applications.

It follows, from the properties considered, that one and the same code set can be represented by several different type A code rings. On the basis of the topological method developed by der Bruyn [2], it is possible to compute the number of different code rings:

$$p = 2^{2^{n-1}-n}.$$

The fact that a type A code ring can assume several values is indicative of the presence in it of nonrealized connections. Therefore, by giving up the multivaluedness, one can probably shorten the ring sequence necessary to express the same code set.

As a means of expressing a code ring by a fewer number of elements, we us the reciprocal substitution property. With this, we require that the following condition hold: the substitution must so change the code ring that the new ring contains none of the codes which entered into the original ring. Thus, a ring of this new type will contain only a portion of the codes. The remaining portion of the codes will enter into a code ring obtained by transforming the original ring. It is clear that, to execute this task, it is necessary to use those substitutions which would affect all the symbols forming the code ring. In particular, it is convenient for this purpose touse cyclical substitutions.

Thus, the problem reduces to the finding of such a partial code ring that, from it, by a strict rule, one may get to other partial code rings, whereby the set of these code rings must represent the complete code set.

As applied to number-positional codes, we prove the following: any code set of number-positional codes can be presented in the form of a grouped ring, in which each group is formed by cyclic substitutions (permutation) of its elements.

For the proof, we denote the operator of cyclic substitution by

$$\alpha = \begin{pmatrix} abc \ldots ik \\ bcd \ldots ka \end{pmatrix},$$

where a,b,c,...,k are the symbols forming the code combinations.

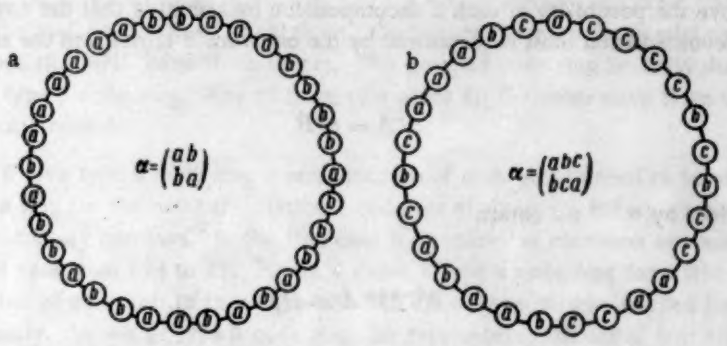


Fig. 3. Type B code rings for number-positional codes. In a, k = 2, n = 6, k^n = 64, m = 32. In b, k = 3, n = 4, k^n = 81. m = 27.

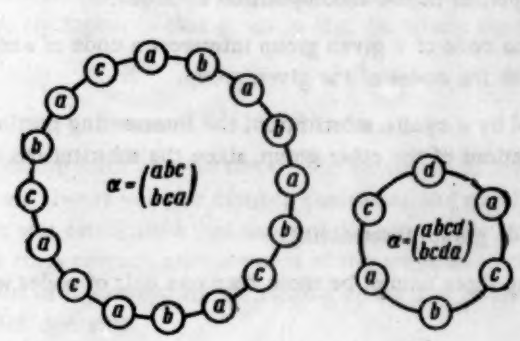


Fig. 4 **Fig. 5**

Fig. 4. Type B code ring for a quality-shift code. k = 3, n = k, $k(k-1)^{n-1}$ = 48, m = 16.

Fig. 5. Type B code ring for combinations of four symbols taken three at a time. k = 4, $A_k^{k-1} = k! = 24$, m = 6.

The ring obtained by a cyclical substitution of the initial ring will differ from it in that, instead of the first symbol, the second symbol occurs, instead of the second, the third, etc. The first symbol occurs instead of the last. It is clear that, by means of cylical substitutions, we can obtain only k-1 new combinations which are not identical with the original ones. Use of the operator a greater number of times leads to the repetition of the previously obtained ring combinations.

We now decompose all the codes into groups, each group consisting only of those codes which be obtained, one from another, by a cyclic substitution of the symbols. With this, we shall choose for each successive group only such codes as have not already appeared in earlier groups. If we denote the initial codes of the various groups by A, B, C,...,L, we can express all the remaining codes in the form of products of the initial codes by the corresponding operators:

$$\begin{array}{ccccc} A & B & C & \ldots & L \\ \alpha A & \alpha B & \alpha C & \ldots & \alpha L \\ \alpha^2 A & \alpha^2 B & \alpha^2 C & \ldots & \alpha^2 L \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha^{k-1} A & \alpha^{k-1} B & \alpha^{k-1} C & \ldots & \alpha^{k-1} L \end{array}$$

We note that, with this, $A = \alpha^k A$, $\alpha A = \alpha^{k+1} A$, etc. Since only k nonrepeated codes can be found in each group, the number of groups will equal $k^n/k = k^{n-1}$.

It is easy to prove the possibility of such a decomposition by assuming that the contrary is true. The impossibility of the decomposition must be expressed by the occurrence of one and the same code in different groups, for example

$$\alpha^r A = \alpha^s B.$$

By multiplying both sides by α^{k-s} we obtain

$$\alpha^{k-s+r} A = B,$$

which contradicts the assumption of decomposability into groups. Thus, the decomposition into groups is possible.

We now consider the properties of the decomposition obtained.

Property 1. If at least one code of a given group intersects a code of another group, then all the codes of this latter group intersect with the codes of the given group.

Indeed, when transformed by a cyclic substitution, the intersecting portion of the code will be identical with the corresponding combinations of the other group, since the substitutions are made in the latter in accordance with the same rule.

We agree to call such code groups intersecting.

Property 2. Inside a group there cannot be more than one pair of codes which intersect in a given combination of symbols.

Each symbol in a given vertical column (if all the codes of one group are written in columns) can appear only once, in correspondence with the law of formation of the group. Consequently, a given combination of elements cannot appear more than once on the left, in particular, a combination identically equal to a given combination of right elements.

We now prove that any code in a group has an external intersection. In fact, to a given code there correspond $\underline{k}$ codes which intersect it (Cf. the analysis of the complete number-positional set as to the presence of intersections), of which, according to property 2, only one can lie inside the group. Consequently, the number of external intersections can be decreased only to the quantity k-1. For this latter, the inequality $k-1 \geq 1$ is always valid, since $k \geq 2$. This means that for any code and, consequently, in accordance with property 1, for any group of codes, there exists an external intersection. The latter is a necessary condition for the formation of a ring from groups.

The other necessary condition for forming a continuous series of groups is, as before, the impossibility of decomposing it into independent parts. Both these conditions always hold for number-positional codes, so that it may be asserted that this set can always be presented in the form of a grouped code ring. This ring will consist of $\underline{k}$ chains (layers). Under definite conditions, these code chains can be separated, as a result of which the grouped ring will be given in the form of $\underline{k}$ different partial rings. Each of these rings may be obtained from any of the others by a cyclic substitution of symbols. Therefore, to represent a complete code set, it suffices to reproduce any such partial ring and to provide the technical capability of carrying out cyclical substitutions. The condition for separability of code chains is that they be so constructed that their intersections are not realized. In the overwhelming majority of case, separable construction of chains is completely possible. This consequency, of practical value, may be formulated as follows.

By the use of cyclical substitutions, a complete code set may be presented in the form of a partial code ring, the number of whose elements is decreased by a factor of kn in comparison with the number of elements necessary to present the same code set in the form of unconnected combinations.

We shall give the designation of a type B code ring to a ring $\underline{k}$ times smaller than a type B code ring and with the property of reproducing, by means of cyclical substitutions, all possible codes of the given code set.

Figure 2 gives an example of a grouped code ring for a four-place quality-shift code. The cross-hatched circles denote quality a, the slashed circles quality b, and the empty circles, quality c. The number of symbols (qualities) from which the codes are formed equal three. Consequently, each group will contain three codes. As a result, the grouped code ring will have three layers. The grouped code ring is easily divided into three. Any of the partial rings is a type B code ring. Any of these ring gives all the other code rings upon multiplication by the cyclical substitution operator.

Figures 3, 4, and 5 give type B code ring representations of code sets formed in accordance with various laws. Figure 3a, gives a ring for the number-positional code set of six-place binary numbers, while Fig. 3b, gives one for four-place ternary numbers. In the first case the number of elements necessary is reduced from 384 to 32, in the second case from 324 to 27. Figure 4 shows a type B code ring for a five-place quality-shift code ($k = 3$). The number of elements in it is 16, while 240 elements would be needed for the unconnected combinations. And, finally, the use of type B code rings for representing the set of four elements taken three at a time (Fig. 5) gives a decrease from 72 elements to six.

In this paper we did not consider the techniques of transforming code rings. We note only that, from the practical point of view, the simplest transformations may turn out to be the most useful. In this sense, attention should be given to code rings, analogous to that given in Fig. 3a, where the cyclical substitution becomes a simple transposition of symbols.

SUMMARY

Code sets can be presented in the form of code rings. The existence of the latter is proven by the holding of two conditions. These conditions always hold for number-positional and quality-shift codes, and also for codes formed by combinations. It was established that one and the same code set can be represented by various types of code rings. Type A code rings contain all the codes of the set in explicit form. Type B code rings contain in explicit form only a part of the codes; the remaining codes may be obtained by multiplying the initial ring by the cyclical substitution operator.

The ring method of representing code sets allows a significant reduction in the number of elements with which the given set is ordinarily represented. Thanks to this, it is possible to increase capacity proportionally, or increase the accuracy of some coding devices.

The effectiveness of the method is increased by the use of multi-place codes. For smaller code sets, the effect obtained is correspondingly less significant.

LITERATURE CITED

[1] A. N. Radchenko, "Code rings and their use in remote control devices," [in Russian]. Automation and Remote Control (USSR) 18, 8 (1957).

[2] N. G. der Bruyn, A Combinatorial Problem, Koninklijke Nederlandshe Akademie van Wetenschappen, XLIX, 1946.

Received January 9, 1959

ON INCREASING THE PRESSURE OF THE WORKING AGENT IN JET AMPLIFIERS

B. D. Kosharskii

(Khar'kov)

For a long time, both in the Soviet Union and beyond its borders, hydraulic jet regulators were used operated with pressures of the workingagent (transformer oil) not higher than 6 to 8 kilogram/cm². Due to the low value of delivery potential, the radius of action of these regulators was very limited, both horizontally and, particularly, vertically.

The theoretical and experimental analysis of the operation of jet amplifiers, given in [1], made it possible to take the first steps in the direction of broadening the region of applicability of jet regulators. Today, industry produces two-stage jet-valve amplifiers in which the pressure of working agent in the first (jet) stage of the amplifier, p_{js} = 5 to 8 kilograms/cm², is increased to a pressure of p_{vs} = 12 to 15 kilograms/cm² in the second (valve) stage. With this, it has been experimentally established that, if necessary, the working agent's pressure in the second stage can be increased to 25 kilograms/cm² with existing amplifiers. The technical solution adopted allowed the radii of action of these regulators to be increased to 150 to 200 meters horizontally, and to 35 to 40 meters vertically, making them more competitive with other types of regulating devices.

However, certain factors inhibit further increases in the delivery potential of the working agent.

In the existing designs of jet pipes, with pressures of the working agent of p_{js} = 6 to 8 kilograms/cm², the value of the Reynold's number for the flow moving in the rectilinear parts of the pipe reaches Re = 2100 to 2200. If the pressure p_{js} is increased to 9 to 12 kilograms/cm², the Reynolds number enters the critical zone, with Re_{cr} = 2320 to 2700. This leads inevitably to an increase in the pulsation flow, which can lead to undamped oscillations of the jet pipe, which has actually been observed.

Tests of two-stage amplifiers with pressures of p_{vs} = 12 to 15 kilograms/cm² and p_{js} = 5 kilograms/cm² showed that the valve begins to follow the jet pipe if its nozzle tip is displaced δ_n = 0.017 to 0.20 mm, which corresponds to a pressure drop at the leading plunger of the valve device of $\Delta p_{vs} \approx 0.035$ kilograms/cm² and an adjustment force of $p_{vs} \approx 0.30$ kilograms.

With pressure increased to p_{vs} = 25 kilograms/cm² in the second stage of p_{js} = 8 kilograms/cm² in the first stage, there is a corresponding increase in the force necessary to move the valve device up to $p_{vs} \approx 0.50$ kilograms, in connection with some increase in the unbalanced pressure in the radial gaps between the value and the frame of the amplifier.

Up to now it has been assumed that the pressure of the working agent applied to the jet pipe cannot be raised above p_{js} = 8 kilograms/cm²; this also provided an upper limit in practice to the pressure in the amplifier's second stage, p_{vs} = 25 kilograms/cm².

The computed values of the parameters characterizing the operation of an amplifier with a vertical jet pipe (produced in the "Teplo avtomat" factory) are shown, for various pressures at the input, in the nomogram (Fig. 1). It follows from the curves given that, for pressures at the first stage's input of $p_{js} \geq 15$ kilograms/cm², the autooscillations of the jet pipe must be sharply decreased. Experiments confirmed the correctness of this conclusion, and showed that, for p_{js} = 15 to 25 kilograms/cm², the autooscillations of the jet pipe were practically completely absent.*

* With a carefully executed conical nozzle.

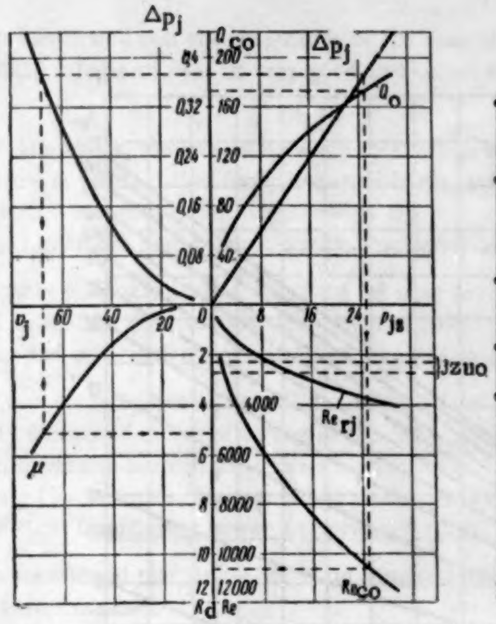


Fig. 1

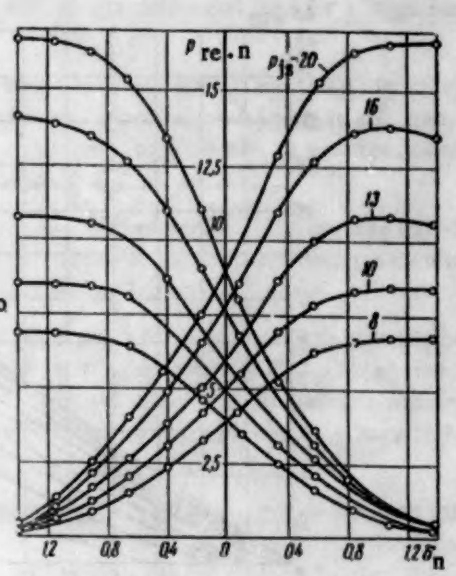


Fig. 2

Fig. 1. Parameters of the vertical jet pipe of factory "Teploavtomata". p_{js} is the pressure before the jet pipe, in kilograms/cm², Re_{c0} is the Reynold's number in the jet pipe's conical nozzle, Re_{rj} is the Reynold's number in the rectilinear portion of the jet pipe, ZUO is the zone of undamped oscillation, Q_0 is the oil discharge through the jet pipe, in cm³/second, Δp_j is the pressure loss in the jet pipe, in kilograms/cm², R_d is the insensitivity of the jet pipe in grams.

Fig. 2. Experimental curves of the pressure variations at the receiving nozzles with various values of working agent pressure at the jet amplifier's input. δ_n is the deviation of the jet pipe from its mean position, in mm, $p_{re.n}$ is the pressure at the receiving nozzles, in kilograms/cm².

When it is considered that the accuracy with which the valve element follows the jet pipe must be within the limits, stated above, of $\delta_n = 0.02$ mm, then the valve drag force p_{vs} must equal 2 to 3 kilograms for p_{js} = = 25 to 30 kilograms/cm².

The values of p_{vs} obtained allow the question of increasing the pressure of the working agent in the amplifier's second stage to 100 to 120 kilograms/cm² to be raised. With this, certainly, a way must be determined of changing the construction of the valve device since, with increased pressure, the jet reaction at the valve openings is sharply increased, as is the inertial axial force, allowing autooscillation to develop in the valve itself. There must also be taken into account the effect of obliteration, which lowers the sensitivity of the valve device when the pressure of the working agent is raised.

In choosing the design parameters of a two-stage amplifier operating with increased pressures of the working agent, it is necessary to take into account the load on the bearings of the jet pipe which determines the sensitivity of the entire device.

The load on the bearings from the force acting along the axis of the jet pipe may be determined from the formula

$$R_0 = \left(p_{js} - \sum_{i=1}^{i=3} \Delta p_i - \frac{\Delta p_{co}}{2}\right)(F_{pi} - F_s) - (p_{js} - \Delta p_i) F_{pi} + \frac{\gamma_o}{g} F_{co} \cos\left(90° - \frac{\theta}{2}\right)\left(\frac{v_{pi} - v_n}{2}\right)^2 + G_{jp}, \qquad (1)$$

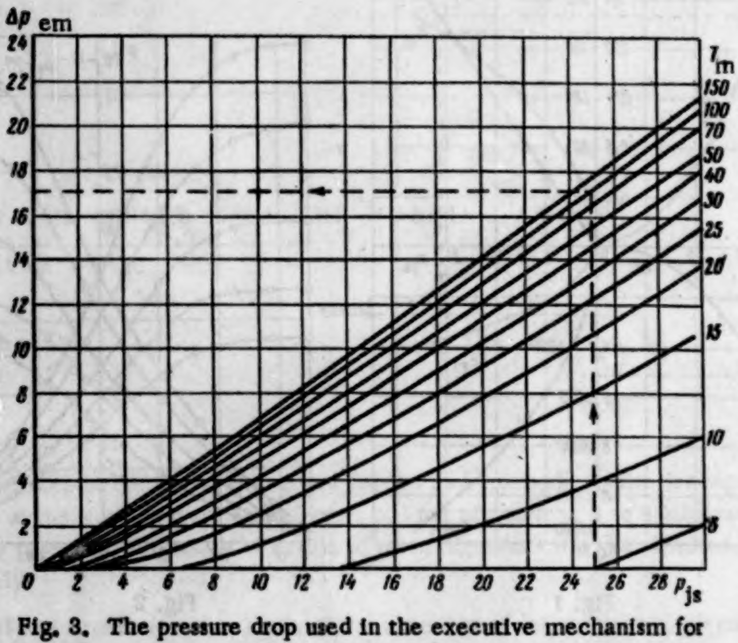


Fig. 3. The pressure drop used in the executive mechanism for various parameters of the jet regulator (for an executive mechanism with a 200 mm travel and plunger diameter of 800 mm). Δp_{em} is the pressure drop used in the executive mechanism, in kilograms/cm² and T_m is the time of complete travel of the executive mechanism, in seconds.

where Δp_1, Δp_2, Δp_3 are the pressure losses at, respectively, the input to the T-pipe, at the bend between the inputs to the jet pipe and in its rectilinear portions, Δp_{c0} is the pressure loss in the conical nozzle, F_{pl} is the area of the straight-through cross section of the pipe, F_s is the area of the output cross section of the conical nozzle, γ_0 is the specific weight of the oil, G_{jp} is the weight of the jet pipe, F_{c0} is the interior surface area of the conical nozzle, θ is the nozzle's angle of taper, v_{pl} is the velocity of flow in the rectilinear portion of the jet pipe and v_n is the velocity of flow at the nozzle's output.

With the inside diameter of the pipe d_{pl} = 5 mm, the diameter of the conical nozzle's output orifice d_n = 1.8 mm and θ = 6 degrees, 18 minutes, we obtain

$$R_0 = 10^{-3} \cdot 0{,}357\, v_n^2 + G_{jp} - 0{,}026\,(p_{js} - \Delta p_1) - 0{,}170\,(\Delta p_2 + \Delta p_3 + 0{,}5\,\Delta p_{co}). \tag{2}$$

In formula (2), v_n is in meters/second, G_{jp} is in kilograms and the pressures are in kilograms/cm².

If the value of R_0 obtained is negative, the force will be applied to the bearings from below upwards. As the velocity of flow of the jet from the nozzle becomes greater, there is an increase in the downward directed component of R_0.

The drag for the jet pipe design being considered is determined from the expression

$$R_a = \mu R_0 \frac{d_0}{2l_{pl}} = 0{,}006\, R_0, \tag{3}$$

where μ is the frictional coefficient at the bearings, taken equal to 0.1, d_0 is the diameter of the journal bearings of the jet pipe, equal to 0.6 cm, l_{pl} is the distance from the axis of rotation of the jet pipe to the point of application of the force from the sensitive element, equal to 5 cm.

It should be mentioned that the magnitude of the drag of the unloaded jet pipe, for p_{js} = 8 kilograms/cm², equal to $R_d \approx 0.0015$ kilograms, was in very good agreement with the experimental data for a large number of jet amplifiers.

For p_{js} = 30 kilograms, the size of R_d was also comparatively low, and did not exceed 0.006 kilograms (Cf. Fig. 1). In view of the fact that force imparted by the sensing element to the jet pipe is ordinarily within the limits of 0.2 to 2.0 kilograms, the magnitude of the drag on the jet amplifier can, in practice, reach as high as 0.02 kilograms, i.e., more than three times the computed value.

It is completely understandable that, if the jet pipe bearings are executed by self-adjusting ball-bearings, the magnitude of R_d will be significantly decreased, and that the effects of the reaction forces, which might arise from bending and other defects of amplifier preparation and assembly, are eliminated.

The use of two-stage hydraulic amplifiers with working agent pressures of p_{vs} = 100 to 120 kilograms/cm² gives a practically unlimited radius of action of this type of regulator in the conditions of ordinary industrial usage. Moreover, hydraulic amplifiers with such pressures of the working agent can be used in a number of new branches of industry (for example, for regulating pressure devices in rolling mills) where they have hitherto gone unused because of their insufficient power and speed of action of the executive mechanisms.

It might be mentioned that single-stage jet amplifiers working with pressures of p_{js} = 25 to 30 kilograms/cm² can also be used independently.

Such amplifiers are simpler and more reliable than two-stage ones, and can successfully operate by using the working medium existing in the regulated object (for example, water). In this case, it is possible to join the regulator organically to the executive mechanisms and to remove from the unit the subsidiary elements used for remote control, as this is done, for example, with steam turbines.

Figure 2 shows the experimentally obtained static characteristics of randomly selected jet amplifier units working with increased pressures of the working agent.

On the basis of the formulae given in [1]* and also in [2, 3], one may determine the relationship between the pressure drop used in the executive mechanism, $\Delta p_{re.n}$ and the time of the mechanism' total travel, T_m, for various values of p_{js} (Fig. 3).

The magnitude of $\Delta p_{re.n}$ is defined by the expression

$$\Delta p_{re.n} = 0{,}92 \frac{\gamma_0}{g F_{re.n}} \left[\sum_{i=0}^{i=k} \Delta F_{re.i} (v_{rj.i} - k_p v_{em})^2 - \right. \qquad (4)$$

$$\left. - \sum_{j=m}^{j=s} \Delta F_{re\,j} (v_{rj.j} + k_p v_{em}{}^2 \right],$$

where $F_{re.n}$ is the area of the input cross section of the receiving nozzle, $\Delta F_{re.i}$ and $\Delta F_{re.j}$ are the elementary areas of the receiving nozzles (i is for nozzles, completely covered by the jet, j for partially covered nozzles),** in cm², $v_{rj.i}$ and $v_{rj.j}$ are the mean jet velocities at the areas $\Delta F_{re.i}$ and $\Delta F_{re.j}$, in cm/second, v_{em} is the translational velocity of the executive mechanism's piston, in cm/second, $k_p = F_{em}/F_{re.n}$ = const (the effect of the size of the slope of the jet pipe with respect to the receiving nozzle is ignored), F_{em} is the piston area of the executive mechanism, in cm².

The coefficient 0.92 determines the lowering of magnitude of the jet's frontal pressure due to the eddying arising at the edges of the receiving nozzle and places where the jet spills.

It should be remembered that, in case when translational speed of the executive mechanism's plunger must not be too great, it may be more economical to use a single' stage amplifier, with input pressures up to 25 to 30 kilograms/cm², than a two-stage amplifier with analogous pressures in the second stage.

* The formula given in [1] were partially changed so that a special determination of the area of the turbulent zone was not necessitated.

** For details on this question, Cf. [1].

By their characteristics, hydraulic jet amplifiers belong which those devices which can operate with quite a wide spectrum of input signal frequencies, up to 20 to 30 cycles [4, 5]. Moreover, as demonstrated by tests on special stands, these amplifiers are completely vibration – and shock-stable.

Therefore, both single-stage and two-stage jet amplifiers with increased working agent parameters, in conjunction with electromechanical and other types of input elements, can be widely used in the future for the purposes of control and following, particularly in those cases when signifant power, rapid speed and a large radius of action are required.

LITERATURE CITED

[1] B. D. Kosharskii, "Certain questions in the design of hydraulic jet amplifiers," [in Russian]. Automation and Remote Control (USSR) 17, 7 (1956).

[2] G. N. Abramovich. Turbulence-Free Liquid and Gas Jets. [in Russian] . Gosénergoizdat (1948).

[3] M. Ya. Alferov. Hydromechanics. [in Russian]. Izdatelstvo Ministerstva Rechnovo Flota (1952).

[4] L. M. Krassov and N. P. Koslov, "Hydraulic vibration contours," [in Russian]. Otchet IAT AN SSSR (1956).

[5] M. Z. Litvin-Sedoi. Hydraulic Instruments in Remote Control Systems. [in Russian]. Mashgiz (1956).

Received July 31, 1958

NOMOGRAMS FOR THE ANALYSIS AND SYNTHESIS OF AUTOMATIC STABILIZATION SYSTEMS

P. S. Matveev and V. N. Plotnikov
(Moscow)

Nomograms, which may be used for the analysis and synthesis of automatic stabilization systems, are given for the desirable logarithmic amplitude characteristics.

Today, frequency methods [1-4] based on the use of desirable logarithmic amplitude characteristics are widely used for the synthesis of automatic control systems. Frequency methods for designing stabilization systems are less widely used due to certain specific features of the problem and the absence of methods for defining desirable characteristics. The present paper proposes a method of defining the desirable transfer functions of closed automatic stabilization systems. After the desirable logarithmic characteristics are obtained, the ordinary methods [1-3] may be used for synthesizing the correcting devices.

As dynamic indicators which characterize an automatic stabilization system, we consider (Fig. 1) the following quantities:

1) the magnitude of the maximum deviation, x_{max}, engendered by a unit step-function disturbing stimulus;

2) the duration of the transient response, T*, engendered by the same stimulus;

3) the time, t_r, for the deviation to attain its maximum value.

Below, we consider nomograms which allow the determination of the quality indicators from the form of the logarithmic amplitude characteristic of a closed automatic control system, and also allow the determination of the form of the desirable logarithmic amplitude characteristic fromgiven quality indicators. The nomograms were constructed for systems which are astatic with respect to the disturbing stimulus, since the nomograms developed in [3] can be used for static systems.

In order that an automatic control system be astatic with respect to the disturbing stimulus, the low-frequency asymptote of the desirable logarithmic amplitude characteristic of the closed system must have a slope of 20 decibels/decade, i.e., the system's transfer function must have the factor s in the numerator.

Since any physically realizable system has a limited pass band, it is clear that the high-frequency asymptote must have a slope of −20n decibels/decade, where n can take the values of 1,2,3,...

To obtain the simplest transfer function of the correcting device, the high-frequency asymptote of the desirable logarithmic amplitude characteristic must differ as little as possible from the object's logarithmic amplitude characteristic.

* For the duration of the transient response, we take the least time after which the controlled quantity gives $|x(t)| \leq \Delta$. For computing the nomograms, the magnitude of Δ was taken equal to 5% of the unit step-function disturbing impulse which engendered the deviation of the controlled quantity, x(t).

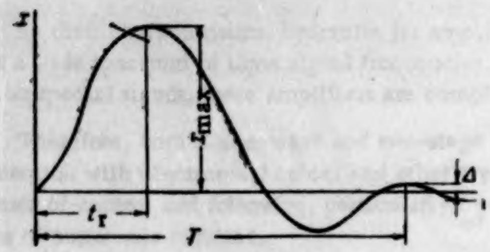


Fig. 1

As for the medium-frequency portion of the desirable logarithmic amplitude characteristic of the closed system, analysis of actual automatic control systems shows that, depending on the quality requirements, it may equally well (Fig. 2): have a slope of 0 decibels/ / decade, −20 decibels/decade, −40 decibels/decade, 0 and 20 decibels/decade, 0 and 40 decibels/decade, −20 and 40 decibels/decade.

Starting from what has already been said, we assume that the transfer function of the closed automatic control system has the form

$$Y_1(s) = \frac{ks}{(T_1s+1)(T_2s+1)(T_3s+1)}. \tag{1}$$

In the particular cases when, in (1), we have $T_1 = T_2$, $T_2 = T_3$, $T_1 = T_2 = T_3$, $T_3 = 0$, $T_1 = T_2$, and $T_3 = 0$, the transfer function has the form:

$$Y_2(s) = \frac{ks}{(T_1s+1)^2(T_3s+1)}, \quad Y_3(s) = \frac{ks}{(T_1s+1)(T_2s+1)^2},$$

$$Y_4(s) = \frac{ks}{(T_1s+1)^3}, \quad Y_5(s) = \frac{ks}{(T_1s+1)(T_2s+1)}, \quad Y_6(s) = \frac{ks}{(T_1s+1)^2}. \tag{2}$$

The asymptotic logarithmic amplitude characteristics of the closed systems with the transfer functions in (1) and (2), are shown in Fig. 2 (characteristics 1-6). As one of the parameters in constructing the nomograms in Figs. 3a, 3b, 3c and 3d, we chose the magnitude of L_1, which is the value of the ordinate of the asymptotic logarithmic amplitude characteristic for $\omega = \omega_1$.

Here, ω_1 ω_m, ω_2/ω_m and ω_3/ω_m are the relative mating frequencies, $\omega_1 = 1/T_1$, $\omega_2 = 1/T_2$ and $\omega_3 = 1/T_3$.

The nomograms (Figs. 3a, 3b, 3c, 3d) were constructed for the following fixed values of L_1: 5, 0, −5 and −10 decibels. The graphs show the dependencies of the dynamic indicators considered, $T\omega_1$, $t_r\omega_1$ and x_{max}, on the relative mating frequency ω_3/ω_1 for fixed values of $\omega_2/\omega_1 = 1.0$ or 2.0.

Thus, to determine the desired quality indicators, it is first necessary to find ω_1, L_1, ω_3/ω_1 and ω_2/ω_1 from the given logarithmic characteristic.

On the nomograms of Figs. 3a, 3b, 3c and 3d, the curves for $T\omega_1$ are solid lines, the curves for $t_r\omega_1$ are dashed lines and the curves for x_{max} are alternating dots and dashes. If the values of the parameters ω_2/ω_1 and L_1 differ from those for which the nomograms were constructed, the quality indicators can be determined by interpolation. An example will show the use of the nomograms.

Let $L_1 = 0$, ω_2 $\omega_1 = 2$, ω_3 $\omega_1 = 4$, $\omega_1 = 1$. Then, on the basis of Fig. 3b, the desired dynamic indicators will be

$$T = 9.6 \text{ sec}., \quad t_r = 1.3 \text{ sec}., \quad x_{max} = 0.48.$$

If $\omega_1 =$ 1, for example, if $\omega_1 = 0.4$, then

$$T\omega_1 = 9.6, \quad t_r\omega_1 = 1.3 \; ; \; T = \frac{9.6}{0.4} = 24 \text{ sec}., \quad t_r = \frac{1.3}{0.4} = 3.25 \text{ sec}.$$

If $\omega_1 = 2$, then

$$T = \frac{9.6}{2} = 4.8 \text{ sec}, \quad t_r = \frac{1.3}{2} = 0.65 \text{ sec}.$$

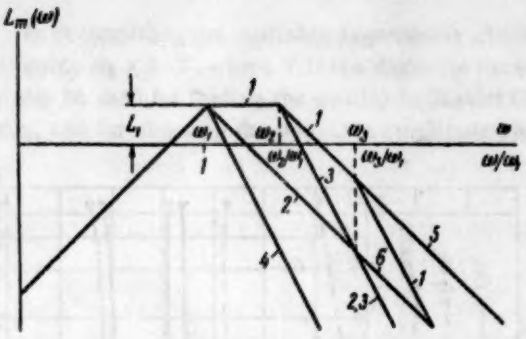


Fig. 2

If the dynamic indicators T and x_{max} are given (in certain cases, t_r may be given) then, by means of the nomograms, the parameters of the desirable logarithmic amplitude characteristics may be chosen. It should be mentioned that the given problem is not single-valued, since one and the same set of quality indicators may be obtained with different logarithmic amplitude characteristics. By means of the nomogram it is also possible to select the desirable logarithmic amplitude characteristic which determines the minimum of any of the quality indicators if the others are given.

Figure 4 gives the family of asymptotic logarithmic amplitude characteristics of a closed system. In the construction of this family, an expression in which x_{max} and T* enter directly was used:

$$Y(s)=\frac{0{,}667x_{max}Ts}{1+\frac{Ts}{2}+\frac{T^2s^2}{10}+\frac{T^3s^3}{120}}\approx\frac{0{,}667x_{max}Ts}{(0{,}2Ts+1)\,[(0{,}2)^2T^2s^2+2\times0{,}7\times0{,}2Ts+1]}. \quad (3)$$

For simplicity, T was taken equal to 5 in the calculations.

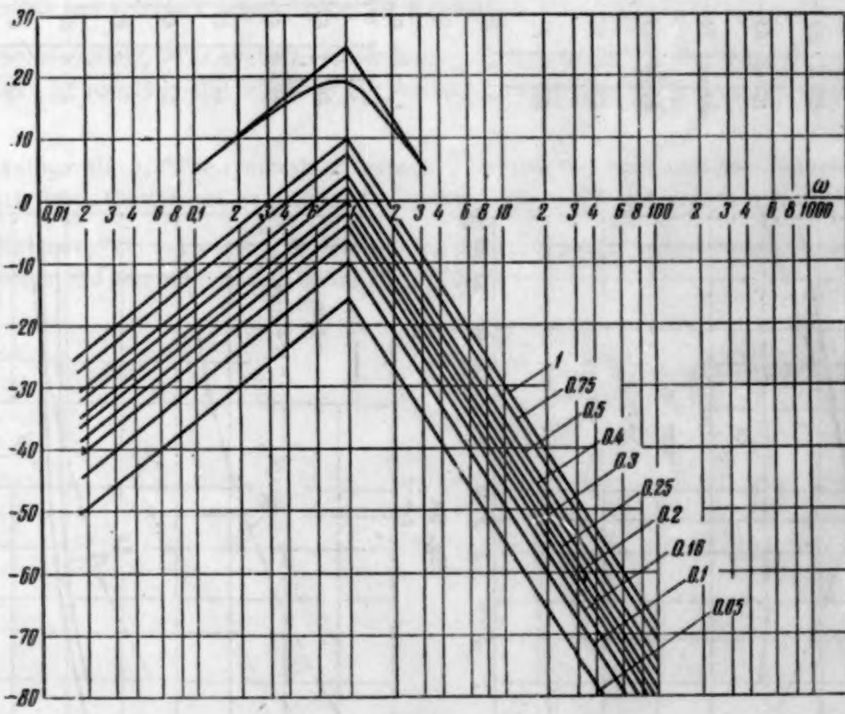


Fig. 4

The characteristics were constructed for values of x_{max} = 1, 0.75, 0.5, 0.4, 0.3, 0.25, 0.2, 0.15, 0.1 and 0.05 (the corresponding logarithmic characteristics are marked by these numbers). The upper part of the figure gives the graph of the correction, δ, which must be added to the asymptotic logarithmic amplitude characteristics to obtain the actual characteristics.

* Formula (3) was obtained with the assumption that the impulsive response of the system was given on a finite interval T by a a polynomial of degree r; the system of coefficients was also assumed to be given. This method [5] may also be used for other cases, for example, for finding the desirable transfer function of static systems.

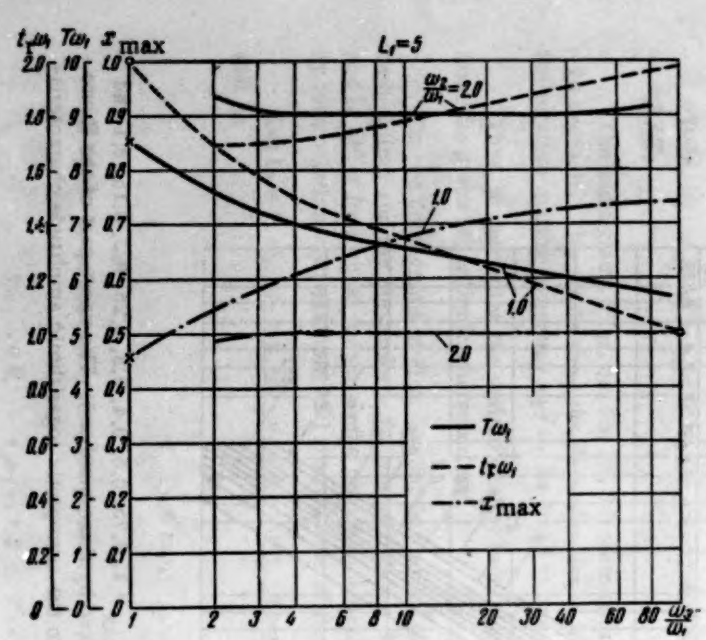


Fig. 3a

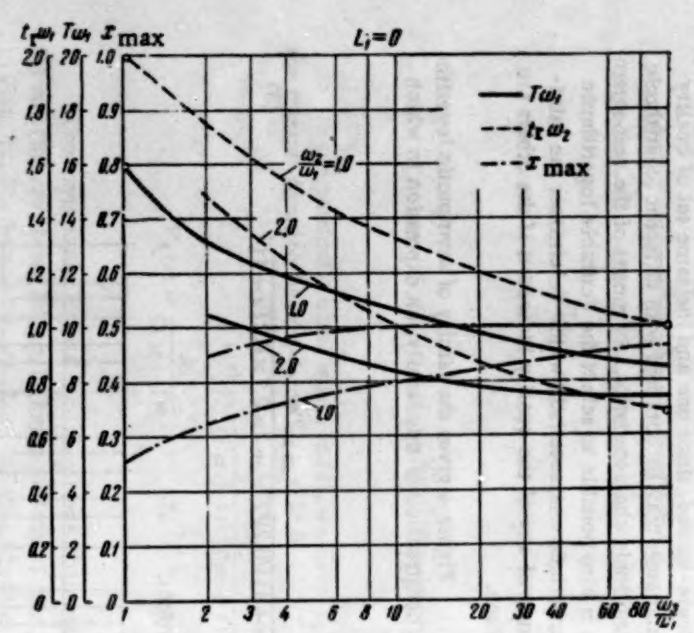


Fig. 3b

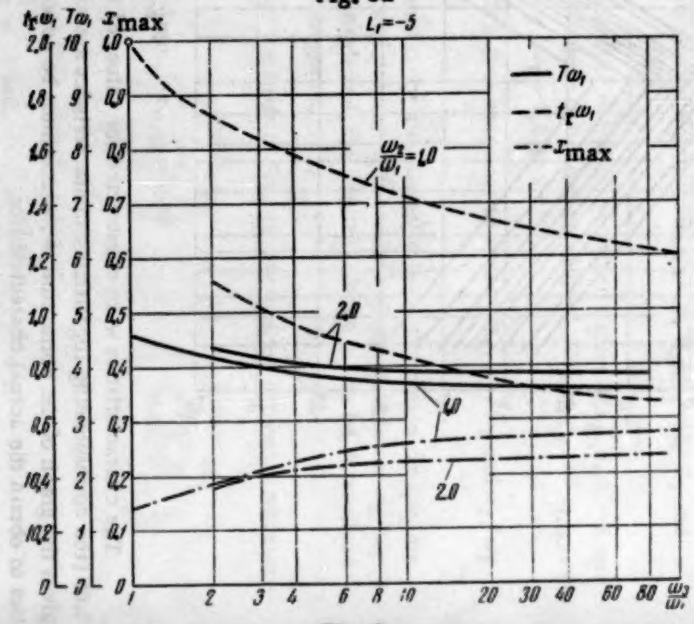


Fig. 3c

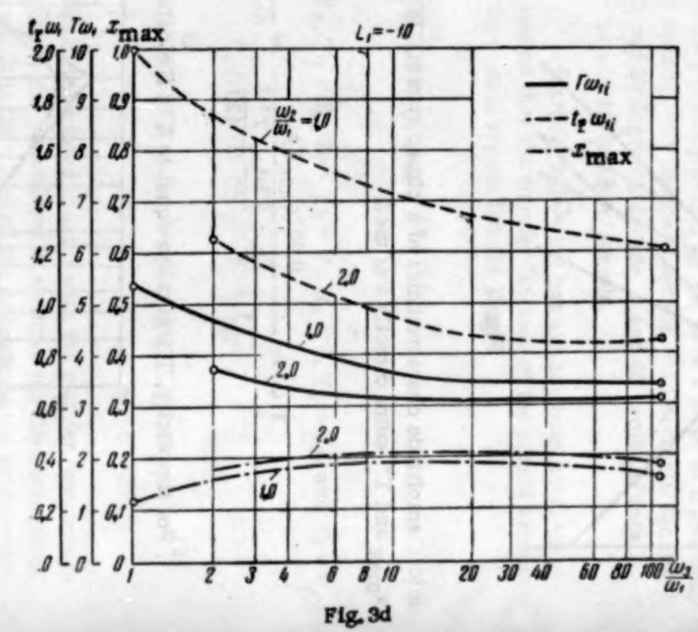


Fig. 3d

In determining the desirable logarithmic characteristic, the mating frequency ω_1 is obtained from the relationship $\omega_1 = 5\ \ T$, where T is the desirable duration of the transient response. The given characteristics may also be used for finding the quality indicators (T and x_{max}) from a concrete logarithmic amplitude characteristic, and for choosing the desirable amplitude characteristics of the closed systems.

SUMMARY

Nomograms were developed for choosing the desirable logarithmic amplitude characteristics of closed automatic stabilization systems.

The nomograms permit a direct determination from the form of the logarithmic amplitude characteristics of a closed system (for minimum-phase type systems) of the following dynamic indicators: The regulation time T, the magnitude of the maximum deviation x_{max} and the time, t_r, for the controlled quantity to reach its maximum deviation.

LITERATURE CITED

[1] Fundamentals of Automatic Control. [in Russian]. Edited by V. V. Solodovnikov. Mashgiz (1954).

[2] V. V. Solodovnikov, "The synthesis of correcting devices of servosystems for standard stimuli," [in Russian]. Automation and Remote Control (USSR) 12, 5 (1951).

[3] V. V. Solodovnikov, "The synthesis of correcting devices for servosystems by means of optimal and standard logarithmic frequency characteristics," [in Russian]. Automation and Remote Control (USSR) 14, 5 (1953).

[4] V. V. Solodovnikov, "The synthesis of correcting devices for automatic control systems," [in Russian]. Trudy II of the All-Union Conference on Automatic Control Theory, volume 1. Izdatelstvo AN SSSR (1955).

[5] P. S. Matveev, "On one method of determining desirable logarithmic frequency characteristics," [in Russian]. Automation and Remote Control (USSR) 18, 1 (1957).

Received July 25, 1958

MAGNETIC-CRYSTAL AMPLIFIERS

O. A. Kossov
(Moscow)

A description is given of a simple circuit for a magnetic-semiconductor amplifier in whose output stage are used transistors operating in the switching mode.

By joining semiconductors and magnetic stages, one can implement circuits with high speeds of action and high gains. As a rule, the input stage is realized by planar transistors (PT) and the output stage is a magnetic amplifier (MA) [1, 2].

The mastering of the industrial production of power PT's allows, in many cases, triodes to be used in amplifiers' output stages. Amplifiers operating as class A or class B possess low efficiency and small utilization coefficients of the triodes,* k_u. Therefore, for significant power to the load (on the order of hundreds of watts), it is advantageous to use the PT switching mode.

For a PT wired in a common emitter circuit (Fig. 1a), the switching mode is characterized by the fact that the triode can remain for long periods of time only at two working points M and N (Fig. 1b). At point M the triode is cut off, the collector and load currents are close to zero and practically all the voltage from the supply is applied to the triode. At point N, the triode conducts, the collector and load currents are high, the voltage drop across the triode is small, and practically all the source voltage is applied to the load. The triode operates at point M if the potential of the base is positive with respect to the emitter and the base current suffices to saturate the triode. The power dissipated by the triode at these working points does not exceed 4% of the maximum power supplied by the triode to the load. The power gain $k_p > 1000$, the triode's utilization coefficient $k_u > 100$.

It is known that, for a class A amplifier, $k_u = 0.5$; for class B operation, $k_u = 2.47$. Consequently, a PT operating in the switching mode can deliver power to the load which is tens of times larger than that of a linear amplifier with the same power dissipation. If the supply source voltage U_s is changed, the load line is displaced parallel to itself. With this, working point M is translated along line OM and point N along line ON.

Use of triodes working in the switching mode in the output stage permits a significant decrease in the dimensions and weight of the amplifier. For controlling switching of the triodes one may use a magnetic amplifier based on cores with rectangular hysteresis loops.

The circuit for such an amplifier, with a dc voltage source, suggested by Collins [3], is shown in Fig. 2. Voltage amplifier VA, constructed of switching triodes, provides a rectangular voltage at its output. The secondary winding of this transformer supplies the magnetic amplifier MA, used for width-pulse modulation of this voltage. The MA load is two switching triodes, T_1 and T_2, whose pass interval is modulated linearly from zero to 50% by means of control current from the MA. In order for the cut off potential to be applied to the triode bases in the intervals of unsaturated core state, and to pass the magnetizing current for the MA, biasing diode D_2 is used, its supply coming from the insulated biasing source E_{bi} via resistor R_{bi}. The amplifier has a nonreversible dc output. The load can be an active, or an inductive-ohmic, impedance. In the latter case, the load must be shunted by diode D (Fig. 2).

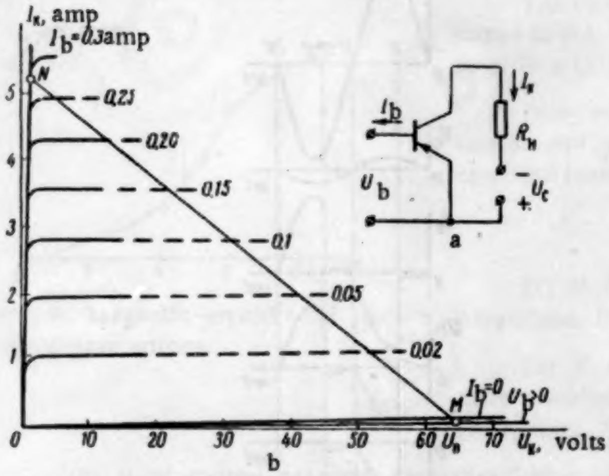


Fig. 1. a) is the triode wiring diagram, b) is the collector's volt-ampere characteristic.

The voltage transformer can be built to operate at higher frequencies. With this, the dimensions and weights of the transformer and the MA will not be large, and the speed of response of the MA will be high. The use of a magnetic amplifier as the input stage allows easy summation of several signals. The disadvantage of the amplifier is the complexity of the scheme which requires an individual voltage transformer for supplying the magnetic amplifier, and a circuit for cutting off the triodes, with an insulated dc source.

In various heat regulators and in controlled low-power resistance furnaces, the amplifier is an active impedance but the supply comes from an ac source. In these and similar cases, a simpler magnetic-crystal amplifier (MCA) circuit, suggested by the author, may be used. In this circuit there is no voltage transformer, and cutting off of the triodes in the intervals of unsaturared MA core states is implemented more simply.

Let us consider the principle of operation of the MCA, whose circuit is shown in Fig. 3a. The output stage is crystal amplifier CA, based on two switching triodes, T_1 and T_2, supplied from the ac source, $U_{n\sim}$, vis diodes D_1 and D_2. The input stage, the magnetic amplifier, is connected to the same ac source via transformer T_n. The load on the magnetic amplifier is winding w_s of transformer T_C, by means of which switching voltage U_s is introduced into the base circuits of triodes T_1 and T_2. Cutoff voltage U_C is applied to winding w_C of transformer T_C from winding w_{III} of transformer T_n. The two secondary windings of T_C introduce the differences of these voltages into the base circuits of the corresponding triodes.

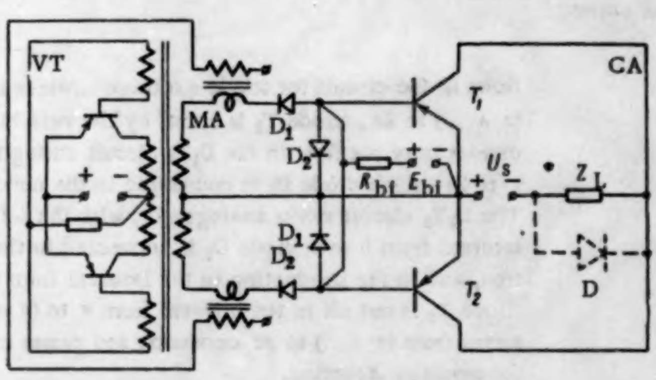


Fig. 2. Circuit of a magnetic-crystal amplifier with a dc voltage supply.

Let the positive half-wave of supply voltage $U_{n\sim}$ correspond to the conducting half-period of diode D_1. The load current will then depend on the conductance of triode T_1. Let the saturation angle of the MA cores equal γ. Then, in the interval from 0 to γ, when the MA cores are not saturated, the cutoff voltage U_c is larger than the switching voltage U_s, due to the flow of magnetizing current in the MA, and the potential, U_{b_1}, of the base of triode T_1 is positive. The triode is cutoff, the load current equals zero. For $\omega_t = \gamma$, one of the MA cores is saturated, U_s increases by a jump, the base potential U_{b_1} becomes negative, and current I_{b_1} flows in the base circuit, saturating the triode. The load current I_L also increases by jumps up to the magnitude determined by the network voltage and the load impedance. In the interval from γ to π, as $U_{n\sim}$ varies, so also does the load current and the triode base current, while complete saturation of the triode is maintained at every point. In the interval from π to $(\pi + \gamma)$, the base of T_1 continues to remain negative and a small current

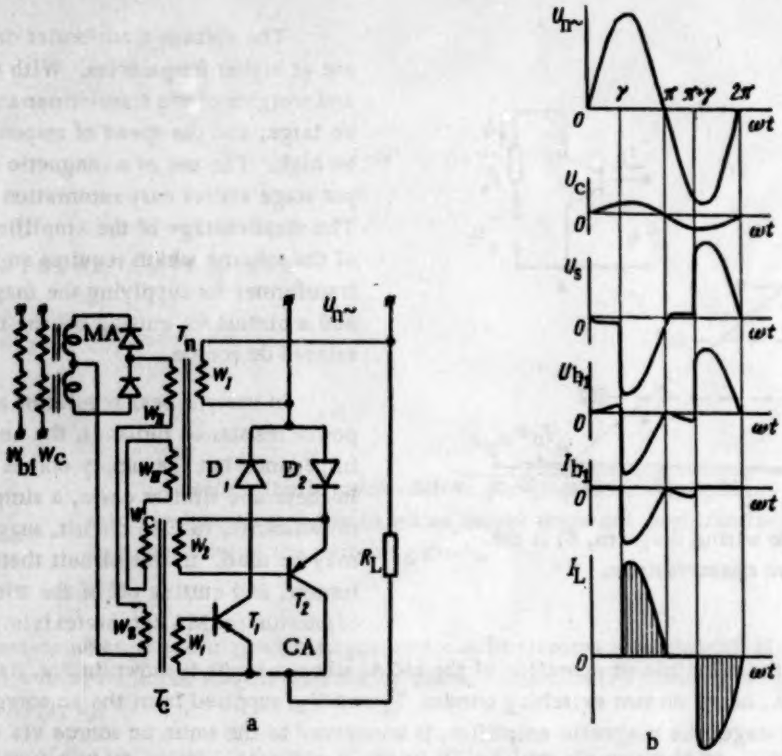


Fig. 3. a) is the circuit of the magnetic-crystal amplifier supplied from an ac source, b) gives the theoretical curves of the variations of the basic vari ables which characterize amplifier operation: U_n is the network voltage, U_c is the cutoff voltage, U_s is the output of the magnetic amplifier, U_{b_1} and I_{b_1} are, respectively, the potential and current of the base of triode T_1, I_L is the load current.

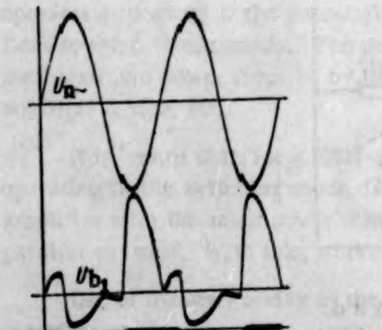


Fig. 4. Oscillograms of voltage and current in the MCA circuit (frequency of 50 cycles).

flows in the circuit for triode excitation, while in the interval from (π +) to 2π, triode T_1 is cutoff by the switching voltage. However, current may not flow in the D_1T_1 circuit during the entire interval from π to 2π, since diode D_1 is connected in the nonconducting direction. The D_2T_2 circuit works analogously, with the sole difference that in the interval from 0 to π diode D_2 is connected in the nonconducting direction, and in the conducting in the interval from π to 2π. With this, triode T_2 is cut off in the interval from π to (π +), and in the interval from (π +) to 2π conducts, and passes current to the load in the opposite direction.

The oscillograms of Fig. 4 gives the curves for U_n , U_{b_1}, I_{b_1} and I_L, which coincide closely with the theoretical curves shown on Fig. 3b.

By regulating the current controlling the angle of saturation of the MA cores, it is possible to vary, within wide limits, the effective value of load current of the magnetic-crystal amplifier.

An experimental specimen amplifier with 170 watt output power, built of two type P4U triodes without additional heat disposal, is characterized by the following data: U_n = 45 volts, f = 50 cycles, R_L = 10.6 ohms, amplifier weight is G = 0.8 kilograms, R_{in} = 15.5 ohms, power gain is $k_p = 1.5 \cdot 10^5$ and efficiency for maximum signal is $\eta \approx 97\%$.

The MA cores and the transformers were made as toroids of material 50 NP.

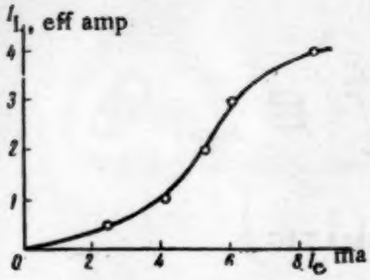


Fig. 5. Magnetic-crystal amplifier characteristics.

The characteristics of the magnetic-crystal amplifier, presented in Fig. 5, are analogous to the characteristics of magnetic amplifiers built of cores with rectangular hysteresis loops.

Based on the principles given above, one can construct nonreversible and reversible amplifier circuits with dc or ac outputs. The amplifier load can be a pure active impedance.

LITERATURE CITED

[1] N. Jasper, J. Taylor and W. White, Transistor Magnetic Amplifiers. Electrical Manufacturing, No. 9, 1957.

[2] V. A. Naidis, A. T. Rozinov and B. Ya. Rozman, "Electric drive machine feeds with magnetic and semiconductor amplifiers," [in Russian]. Stanki i Instrument, 6 (1957).

[3] H. W. Collins, Magnetic Amplifier Control of Switching Transistors. Trans. AIEE, vol. 75, pt. I, p. 585, 1956.

Received December 1, 1958

POLARIZED-RELAY VIBRATION CONTROL BLOCK FOR PNEUMATIC DRIVES

N. S. Gorskaya and B. I. Myzin
(Moscow)

A polarized-relay vibration control block designed to operate with pneumatic drives is considered.

The basic requirements imposed on the vibration control blocks of pneumatic drives, at whose outputs autooscillations are inadmissible, are formulated, the characteristics of the block are given, as well as oscillograms of the transient responses in the individual portions of the control block and the results of testing it in a complete scheme of a servomechanism in conjunction with a pneumatic drive.

In a number of cases in the development of high-quality automatic control systems there arises the problem of constructing fast-acting servomechanisms. Several types of fast-acting electropneumatic servomechanisms were developed at the IAT AN SSSR. V. A. Trapeznikov and V. V. Petrov [1] proposed a fast-acting servomechanism with internal vibration loops tuned to a frequency within the limits of 15 to 30 cycles. The power element of this servomechanism was a pneumatic drive which operated in the autooscillatory mode with an amplitude of 10 to 15% of the full travel.

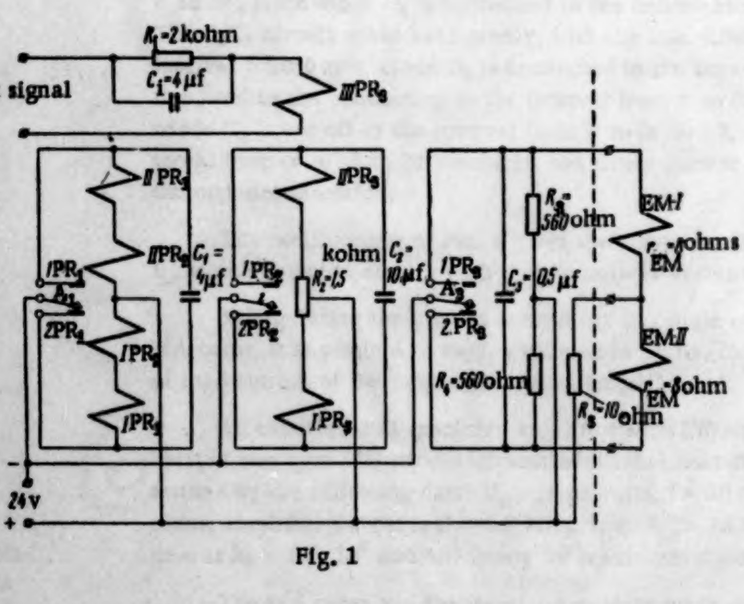


Fig. 1

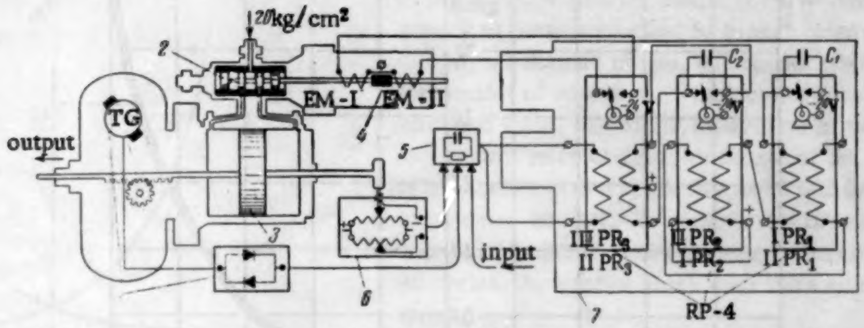


Fig. 2. 1) is the nonlinear velocity feedback; 2) is the valve; 3) is the pneumatic drive's piston; 4) is the electromagnet; 5) is the differentiating loop; 6) is the rigid feedback; 7) is the control block.

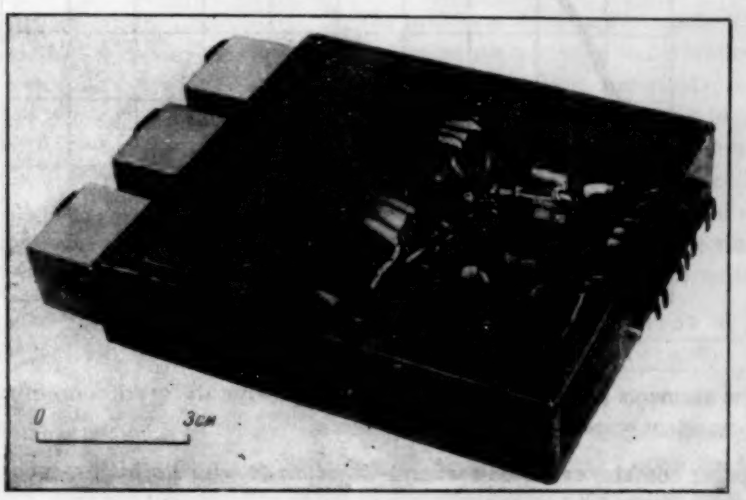


Fig. 3

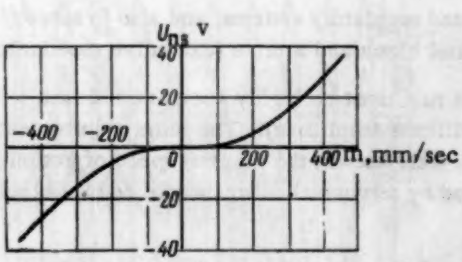


Fig. 4

V. V. Petrov and the authors of the present paper developed an electropneumatic servomechanism with internal vibration blocks at a 50-cycle frequency. Stable operation of the vibration loop was guaranteed by introducing an additional 50-cycle synchroniz tion voltage. This servomechanism had output autooscillations of insignificant amplitude.

The authors developed a vibration control block for electropneumatic servomechanisms, the block being based on polarized relays [2, 3] and designed to operate with modified pneumatic drives.

The development of this block was based on the idea of V. S. Kulebakin regarding "stimulating stabilization" of relay systems by imposing oscillations of increased frequencies from a constant source. This idea was further developed in the works of G. S. Pospelov [4].

A servomechanism with stimulating stabilization, consisting of an autonomous generator forcing 80-cycle oscillations and a nonlinear speed feedback, possesses good dynamic qualities and a complete absence of autooscillations at the output.

There is a brief description of the enumerated servomechanisms in [5]. Also to be found there is a formulation of the basic steps in servomechanism design: choosing the form of the transient response, choosing the

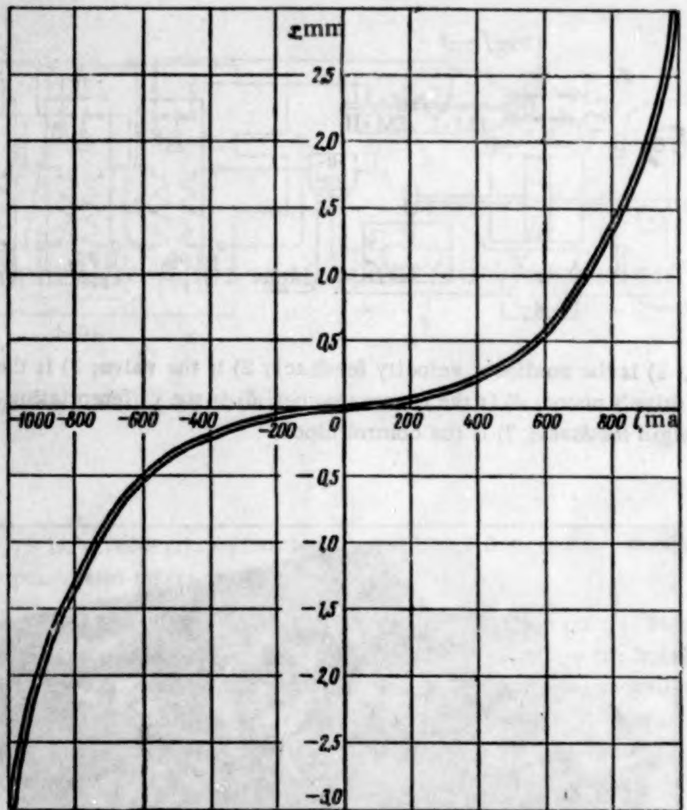


Fig. 5

characteristics of the elements and the connections so as to provide the given form of transient response, and oscillograms of the transient responses are also given there.

The present paper contains extended material having to do with a vibration block with stimulating excitation, realized with polarized relays.

1. Basic Requirements Imposed on the Vibration Control Blocks of Pneumatic Drives at Whose Outputs Autooscillations are Inadmissible

Servomechanisms, widely used today in automatic control and regulatory systems, and also in servo (following, or tracking) systems, consist basically of two links: a control block and a drive (executive mechanism).

Servomechanisms employed as steering machines, etc., as a rule must be highly accurate and fast, with no autooscillation at the output (or with autooscillations of insignificant amplitude). The same requirements are imposed on the control blocks of such servomechanisms. As is well known, the greatest speed of action in conjunction with a stable stationary mode of operation, is possessed by servomechanisms whose control blocks have a relay characteristic with small segments of linearity.

Thanks to the presence of a linear zone in the characteristic of the control block, the servomechanism is stabilized about an equilibrium position and, in connection with the presence of a saturation zone, retains all the properties of a relay system, allowing the greatest speed of action to be obtained.

Since oscillations at the output are inadmissible in the steady state of servomechanisms of the type considered, the vibration control block must have a mode of operation which differs from what is common for vibration loops. With this, the output element of the control block, an electromagnet articulated with the drive valve, must perform a steady-state oscillation with an amplitude lying within the limits of valve covering. When there is an input signal which exceeds a certain definite value, the electromagnet, and the valve joined to it, must perform a complete operating translation, uncovering the valve aperture completely. The frequency of oscillation of the electromagnet-valve, determined by the vibration loop, must be chosen on the basis of the filtering properties of the pneumatic portion in such manner that the valve oscillation not engender oscillations of the pneumatic drive.

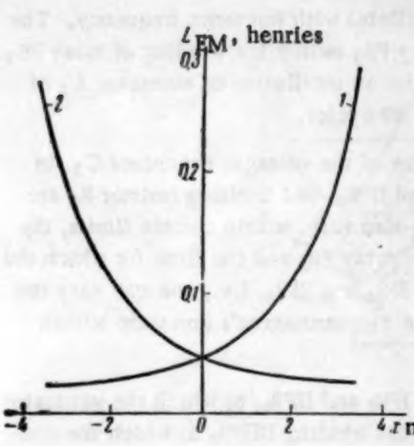


Fig. 6

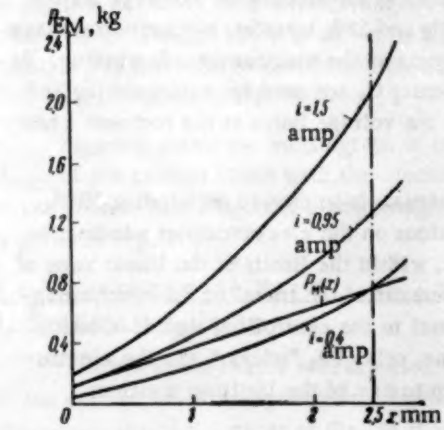


Fig. 7

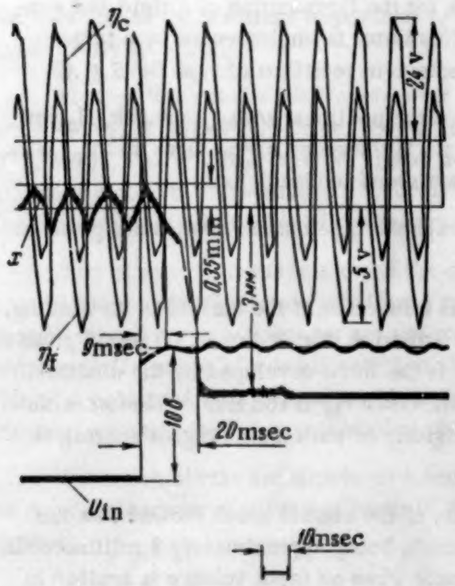


Fig. 8

In order that the electromagnet oscillate with an amplitude lying within the limits of valve covering, its frequency of oscillation must be high.* However, at frequencies higher than 40 cycles, one cannot obtain stable vibratory modes of operation of the control blocks with internal vibration loops, due to the interaction of the electromagnet and control relay circuits, since one of the relay windings is in parallel with a condenser and itself forms a vibration loop (tuned circuit). To remove this interaction and obtain the capability of establishing a frequency higher than 40 cycles, the control block must have an oscillation generator.

This work provides a description of a polarized-relay control block in which the relay characteristic is linearized by means of an oscillation generator of triangular-shaped vibrations.

The nonlinear Z-shaped characteristic of the final amplifier, under the action of the vibrations (forced oscillations) is linearized, or, more accurately, deformed into an uneven nonlinear characteristic which includes linearity and saturation zones. The triangular form was chosen for the curve of the forcing oscillations since it is the best form of those oscillation curves which provide a linear relationship between the average values of the final amplifier's output coordinates and the errors in a narrow range [4].

2. Polarized-Relay Control Block and Its Operation in the Complete Servomechanism Scheme

The block schematic of the vibration control block is shown in Fig. 1, while Fig. 2 gives the block schematic of the servomechanism. An exterior view of the control block is shown in Fig. 3. To eliminate the influence of the final amplifier on the oscillation generator, an intermediate isolating stage with introduced into the block.

The control block consists of three RP-4 polarized relays. Relay PR_1 is in the oscillation generator circuit. The intermediate stage is based on relay PR_2, and relay PR_3 (in the final amplifier) is used to switch in the electromagnet under the stimulus of the forcing and controlling signals.

The triangular oscillation generator is an independent RLC tuned circuit formed by windings IPR_1 and $IIPR_1$ and condenser C_1 in parallel with them. Also included in the tuned circuit are windings IPR_2 and $IIPR_2$ of relay PR_2, connected in series with the corresponding windings of relay PR_1.

The RCL loop is supplied by a 24 volt dc source. The self-excitation pulses are formed periodically by the transfer of the armature of relay PR_1. The frequency of transfer of armature A_1 of relay PR_1 corresponds to 80 cycles. Changing the capacitance of C_1 will change the frequency. Since an ac current of 80-cycle frequency flows through windings IPR_2 and $IIPR_2$ (the same as in windings IPR_1 AND $IIPR_1$),

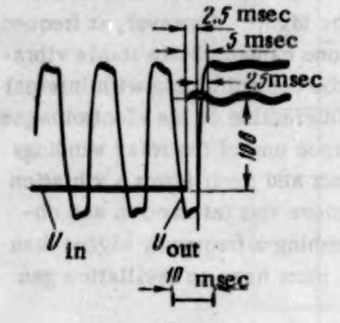


Fig. 9

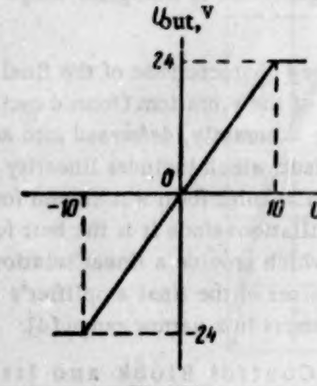


Fig. 10

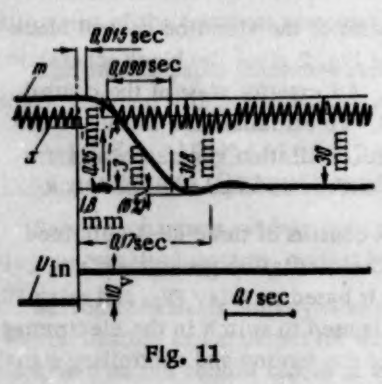


Fig. 11

armature A_2 of relay PR_2 oscillates with the same frequency. The contacts and armature of relay PR_2 switch the winding of relay PR_3. Thus, the tuned circuit provides an oscillation of armature A_3 of relay PR_3 with a frequency of 80 cycles.

For maintaining the form of the voltage, condenser C_2, in parallel with windings IPR_3 and $IIPR_3$, and limiting resistor R_2 are used. By varying R_2, one can also vary, within certain limits, the travel time of the armature of relay PR_3 and the time for which the armature remains at contacts $1PR_3$ and $2PR_3$, i.e., one can vary the amplitude of oscillation of the electromagnet's armature within certain limits.

In addition to windings IPR_3 and $IIPR_3$, to which the generator signals are applied, relay PR_3 has winding $IIIPR_3$ to which the controlling signals are applied. The ratio of the break and make durations of relay contacts $1PR_3$ and $2PR_3$ varies as a function of the amplitude and sign of the controlling signal. The contacts of relay PR_3 are used for switching in the electromagnet winding. At the moment when contacts $1PR_3$ and $2PR_3$ transfer, rectangular voltage pulses with small horns appear at the electromagnet's winding. Resistors R_3 and R_4, and condenser C_3 are used for extinguishing and limiting the magnitude of the voltage horns at the moment when the PR_3 contacts transfer.

When a controlling signals is impressed on winding $IIIPR_3$, the ratio of the pulse durations on the electromagnet windings is changed. Correspondingly, within the limits of the linear zone of the final amplifier's characteristic, the travel of the electromagnet's armature is proportional to the controlling signal. Outside the limits of the linear zone, relay PR_3 "sticks," and the electromagnet proceeds by a jump to one of the limiting positions.

With the given parameters of the vibration block control circuit, the amplitude of the electromagnet's oscillations is $x_a = 0.3$ to 0.4 mm. To provide stability, and also to obtain a desirable form of the transient response for a given speed of action, provision is made in the control block for the introduction of a rigid and nonlinear speed feedback. This latter is implemented by a tachometer generator TG and selenium rectifiers of type 54VS × 45.

The characteristic of the nonlinear speed feedback, $U_{ns}(m)$, is shown in Fig. 4, where m is the speed of the drive and U_{ns} is the voltage in the nonlinear speed feedback path.

The control block was developed for application to the electromagnet whose characteristics are given in Fig. 5-7.

Figure 5 shows the dispacement of the electromagnet armature as a function of the current in its winding. Figure 6 gives the dependence of the electromagnet's induction on the armature displacement. Figure 7 gives the the tractional characteristic of the electromagnet, $P_{EM}(x)$, where P_{EM} is the force developed by the electromagnet, and also the electromagnet's mechanical characteristic, $P_M(x)$, where P_M is the sum of the forces developed by the electromagnet's spring and the frictional forces. The rigidity of the electromagnet's spring is about 300 grams/mm.

Experimental investigation of the transient response in the circuits of the control block showed that the lag of the control block, in conjunction with the electromagnet, was small, being approximately 9 milliseconds. The operation time of the electromagnet, i.e., the time from the moment when an input voltage is applied to

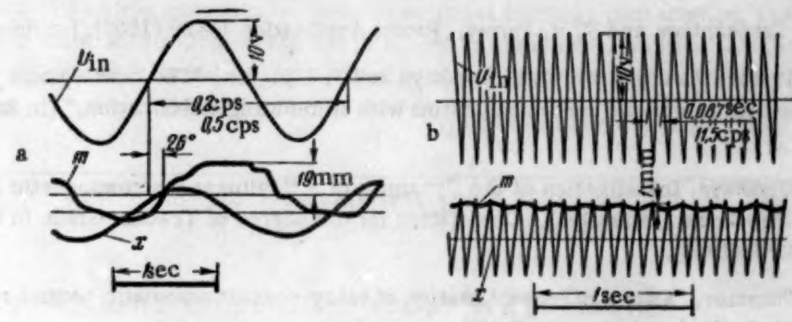


Fig. 12

the control block to the moment when the electromagnet has transferred to its limiting position, equalled 20 milliseconds. The high speed of action of the electromagnet was attained by shunting resistor R_1 by condenser C_1 (to accelerate operation of relay PR_3), and also by introducing supplementary resistors, R_s = 10 ohms, parallel to the two halves of each electromagnet winding (to accelerate electromagnet operation). The transient response of the control block is shown in Fig. 8, where U_{in} is the input signal, x is the electromagnet's displacement, η_c is the voltage oscillation on relay windings I-IIPR$_1$ and I-IIPR$_2$ and η_f is the forced voltage oscillation on relay winding I-IIPR$_3$.

Figure 9 shows the oscillogram of input voltage U_{in} (sticking voltage of relay PR_3) and the output voltage U_{out}, of the control block with the electromagnets disconnected. The oscillogram shows that the total lag time of the control block without electromagnets equals 5 milliseconds. The nonlinear characteristic of $U_{out}(U_{in})$ is given in Fig. 10.

3. Experimental Results of Testing the Joint Operation of the Control Block with a Pneumatic Drive

The pneumatic drive was supplied with air at a pressure of p = 20 kilograms/cm². The transient response in the control block and pneumatic drive, with a nonlinear speed feedback present, is shown in Fig. 11. It is clear from the oscillogram of Fig. 11 that the run-out (overshoot) of the drive does not exceed 6% of maximum travel.

With no input signal, the oscillations of the electromagnet-valve have an amplitude of x_a = 0.35 mm with a frequency of $f_a \approx 80$ cycles. The lag time is t_{lag} = 0.15 seconds; unloaded, and with a travel m = 30 mm, the duration of the transient response is t_S = 0.17 seconds; the maximum unloaded speed is 450 mm/second; the insensitivity is ± 3%.

Figure 12 gives the oscillograms for the external sinusoidal stimuli as operated on by the control block together with the pneumatic drive. For a frequency of 0.5 cycles of the external signals, the phase shift was 26° (Fig. 12a) The greatest external signal frequency passed by the pneumatic drive, as follows from the oscillograms of Fig. 12b, is approximately 11 to 12 cycles. For a frequency of the external signal greater than 12 cycles, the amplitude of oscillation of the pneumatic drive is less than 0.3% of maximum travel.

The experiments made showed the comparatively high dynamic quality of the polarized-relay vibration control block, and also made manifest a number of its superior features in comparison with internal vibration loops. Among these advantages are the stability of operation at high frequency oscillations (80 cycles) of small amplitude (fractions of a millimeter), and also the high speed of response with low weight and small dimensions. Use of a polarized-relay vibration block for pneumatic drive control tends to increase the stability of the servomechanism characteristics as a whole, and increases the reliability of operation of the pneumatic portion.

Vibration blocks are simple to assemble and adjust. Further perfecting of control blocks must be related to the development of vibration loops of the contactless type, which will permit reliability to be increased even further.

LITERATURE CITED

[1] V. A. Trapeznikov and V. V. Petrov. Patent Application 13469 (1952). [in Russian].

[2] B. N. Petrov, V. V. Petrov, N. S. Gorskaya and B. I. Myzin, "The development and experimental investigation of electropneumatic servomechanisms with stimulating stabilization," [in Russian]. Otchet IAT AN SSSR (1954).

[3] N. S. Gorskaya, Investigation of the Dynamics of a Nonlinear Electromagnetic Servomechanism with Nonlinear Speed Feedback. [in Russian]. Dissertation for the degree of Teacher Grade in Engineering Science, part 1, IAT AN SSSR (1955).

[4] G. S. Pospelov, "Stimulating stabilization of relay-contact automatic control systems," [in Russian]. Trydi VVIA im. N. E. Zhukovskovo, 335 (1949).

[5] V. V. Petrov and N. S. Gorskaya, "Rapid-acting electropneumatic drives (electropneumatic servomechanisms)." [in Russian]. Collected papers on the scientific-technological problems of electric-drive automation. Session of the AN SSSR on the Scientific Problems of Automation of Production, 1956. Izdatelstvo AN SSSR (1957).

Received May 30, 1958

Although this article concentrates more on political ideology than science, and expresses views with which we are not in accord or agreement, we have felt that we should publish it in full, in light of our policy of cover-to-cover translation.

Instrument Society of America

CHRONICLE

CYBERNETICS IN THE LIGHT OF LENIN'S BOOK "MATERIALISM AND EMPIRIOCRITICISM"

(Concerning the meeting of the Academic Council of the IAT AN SSSR, devoted to the 50'th anniversary of the publishing of V. I. Lenin's book "Materialism and Empiriocriticism")

On May 14, 1959 in the Institute for Automation and Remote Control of the AN SSSR, there was held a meeting of the Academic Council dedicated to the 50'th anniversary of the publishing of V. I. Lenin's book "Materialism and Empiriocriticism." The participants in the meeting listened delightedly to B. M. Kedrov's paper, "The book by V. I. Lenin,"Materialism and Empiriocriticism" and modern natural science." V. S. Pugachev, A. A. Fel'dbaum and S. M. Shalyutin presented additional papers on the philosophical problems of cybernetics in the light of V. I. Lenin's work of genius.

B. M. Kedrov characterized in detail the value of V. I. Lenin's work for the understanding of those profound processes which began in natural science, particularly in physics, at the turn of this century and which have continued up to the present. The speaker also considered such concepts as "the latest revolution in natural science" and the "crisis in physics" which underlie Lenin's analysis of modern natural science. The great discoveries made in the domain of physical studies on the structure of matter during the half-century since the publishing of V. I. Lenin's book, and that philosophical struggle which unchangingly swirled, and still swirls, about them brilliantly confirm the correctness of the analysis given by Lenin of modern natural science and the outlook for its future development.

In particular, the development of automation, and the branches of physics, mathematics and engineering related to it may be used as an example demonstrating the correction of the conclusions drawn in Lenin's book. All this also related directly to cybernetics.

The opinion is sometimes advanced that the new technological revolution in the Soviet Union will be the practical use of atomic energy as the most powerful energy source. This is not the case. Indeed, it is automation, moving productive processes under their own control, which today plays the same decisive role as was played by power machines in the industrial revolution of the eighteenth century. Cybernetics solves the problem of replacing, not the workman's hand, as was done two hundred years ago, but the brain of the workman, the foreman, the engineer, in production and in other practical domains of human technological activities.

Certainly, no one is speaking of eliminating the brain, i.e., the creative element, from productive processes, but only of the shifting of certain very essential functions, hitherto performed by the brain of the worker or the engineer, to a machine-automation. In the measure that the brain is relieved of these functions which will be transferred to an ever greater degree to cybernetic machines, the human brain (the consciousness of the subject) can reserve to itself the execution of the more complex, more responsible and decisive operations which are not subjected to "automation," being, as it were, the highest control console of the entire process as whole, including therein the cybernetic machines themselves, and operations including the creation of these machines, their programming and their further perfecting.

In his book, V. I. Lenin showed that the crisis in physics and in all of modern natural science in the capitalistic countries consists of the attempt to draw from the revolution in natural science reactionary philosophical and sociological conclusions.

Precisely this occurred with cybernetics with many reactionary thinkers in the countries of modern capitalism. In their eyes, cybernetic machines, in the ideal, are pictured by the capitalist as means for the complete replacement of the human brain. It is naturally assumed, with this, that such "brains" will not have the tendencies, unacceptable to the capitalist, towards strikes, towards revolutions, etc.

Such, in the final analysis of social realities, is the idea of artificial intelligence, of "robots," etc., so widely diffused in certain foreign popular, and even scientific, works. In them is propagandized the theoretical possibility of solving any problem, not only technological but also social and cognitive, by means of electronic computers, including such problems as determining the laws of modern capitalism and of all human societies, questions of war and peace, etc.

The social direction here is completely evident: it consists of "proofs" of the imaginary needlessness of a science of societies, of the imaginary impossibility of scientific prediction based on a knowledge of the laws of societal development, of the imaginary impotence of Marxism to state the actual path of the historical process. All this is announced by the reactionary sociologues and philosophers of cybernetic matters which, certainly, is the result of a profound hypertrophy of their true meaning and value for modern scientific-technological progress.

The question arises, in connection with this, as to whether cybernetic machines are capable in principle, given sufficient size, complexity and time, to replace the human brain and to carry out cogitation, to be, in the ordinary sense of the words, "thinking" and "creating" machines, i.e., to turn themselves into subjects possessing thought, although artificially created by human thought.

This question is of particular interest in connection with the marked recent tendency in the capitalist countries towards an objective idealism (neo-Thomism, neo-Platonism) among certain strata of the supporters of idealistic philosophy (particularly, neo-positivism) and neo-Machian "physical" idealism, i.e., the subjective-idealistic direction. This phenomenon harmonizes perfectly with the idea of replacing the brain of an individual human (the subject) by a cybernetic machine, beyond which stands a higher Essence ("Essence" with a capital letter, i.e., God), "wisely" directing the general course of experience in the world.

The reactionary philosophical and sociological conclusions, drawn by certain authors from cybernetics and its great successes, in no way undermine the value of this science, since it bears no responsibility for these conclusions.

Of course, it is necessary to take issue with the reactionary conclusions and interpretations with all resoluteness, but it is forbidden to throw out, together with the muddy water of the reactionary Weltanschauung which tries to live parasitically off the accomplishments of science, the living, viable origin, cybernetics itself.

It is interesting to note that in connection with cybernetics, just as with quantum mechanics not too long ago, gnosiological conclusions are drawn which are supposed to be the "new" solution to the question of the relationship of subject to object on the basis of the putative latest accomplishments of science and experimental technology.

In his book, V. I. Lenin clearly shows what is the authentic scientific solution of this gnosiological question: the object really exists, outside of, and independently of, our consciousness (of the subject); the subject (our consciousness) is only the image of the object and, consequently, is something secondary with respect to the object. With this, our sensation is not a partition separating subject from object but, on the contrary, a direct connecting link, thanks to which the energy of an external stimulus crosses over into a fact of consciousness.

Consequently, we have the relationship:

OBJECT.....SUBJECT

The subjective-idealistic interpretation of quantum mechanics in particular consisted of this: that between the subject and the object there was introduced a measuring instrument, thanks to which there was implemented a supposedly continuous relationship between the subject (experimenter observing the microcosm) and the object (the microcosm itself).

In the opinion of the "physical" idealists who draw from quantum mechanics the gnosiological conclusions in the sense already cited, the instrument does not simply aid the observer to detect the microprocesses, which

would be impossible to analyze without it, but the instrument is itself included in the flow of the physical microprocesses and becomes one of their constituents, whereby the instrument "prepares" the object itself (the microcosm). As a result, the object turns out to be directly dependent upon the subject and his instruments, by which is supposedly "proved," "on the basis of" quantum mechanical data, the old idealistic idea that there is no object without a subject, the ideal which was particularly promulgated by the Machist Avenarius whose adherent, A. A. Malinovsky, was attacked by Lenin in "Materialism and Empiriocriticism." By uniting the instrument with the subject and subsuming the action of the instrument under the principle of uncontrollability, the "physical" idealist constructed his type of "instrumental" idealism and advanced the concept of "physical" (i.e., prepared by the instrument) "reality," which supposedly supplanted the materialist concept of objective reality (i.e., matter) so explicitly and convincingly developed by V. I. Lenin in his book.

The result is the gnosiological scheme wherein the instrument, which is actually part of the object, part of the material world, is incorrectly included with the subject with the purpose, by means of the "hauling up" instrument onto the subject, of dissolving in them, at least partially, the microobject itself, depriving it of the marks of independence of the subject and, consequently, the marks of materiality, in Lenin's use of that concept.

The essence of the attempt to interpret cybernetics as if it were evidence for the theoretical possibility of replacing human consciousness by machines is nothing else than the false representation of the subject-object relationship, based on the distorted representation of the role of the instrument, in the given case, the machine-automaton, as in the case of "instrumental" idealism. The difference consists only in this: that earlier gnosiological conclusions were drawn in the spirit of subjective idealism, but now they are drawn in the spirit of objective idealism, and are accompanied by the "liquidation" of the subject himself by dissolution in a technical device created by human art.

The result is a gnosiological scheme in which the role of the subject (individual consciousness) is executed by a machine, i.e., an object prepared artificially, but the spiritual beginning is beyond the limits of the individual human being in the form of some "world spirit," or God:

OBJECT + MACHINE-AUTOMATON = "SUBJECT".

The assumption that a machine, as complicated as you like, can opine, think, feel, etc., similarly to an actual living subject, is a sheer misunderstanding, is the result of confusing mechanical or physical problems with philosophical, gnosiological, problems. No machine, independently of the degree of its complexity or perfection, independently of the properties supplied to it (for example, the capability of spontaneous reproduction of itself by similar machines, etc.) can ever stop being a machine, can ever realize itself its deeds and actions, mentally see beforehand the results of the actions proposed by it, as is done by human beings for both the simplest and most complex cases. And this applies not only to nonliving machines but also to lower living organisms, lacking consciousness and executing acts with corresponding efficacity. The machine-automation has not, and cannot have, any idea of the results of its subsequent acts, but all, without exception, that bears on this result and the "goal" that is placed before the machine is communicated to it, all, down to the last sign, by a human being, i.e., by a subject. The programmed actions of cybernetic machines, whether these have to do with just one machine or a whole chain of machines, even such a chain where one machine produces, "gives birth to," another one similar to it, are given by a human being, by virtue of his conscious activity. Properly speaking, the programming placed in the machine by the human being is nothing other than the ideal idea of the worker (subject) at the beginning of the process of the result which should be obtained at the end of this process, whereby the ideal is translated into the material by the materialization of the machine's data. Consequently, under examination, even the cybernetic machine turns out to be only a particular device wherein the worker's ideal notion of a desirable result of a process may be translated from the worker's head (where it exists only as an ideal) to the machine itself in the form of a program of its actions and is thus materalized at the very beginning of the process.

To put it briefly, cybernetic machines give no new data for the posing of the fundamental gnosiological question of the relationship between subject and object.

The transfer to these machines of some functions previously executed by human brains does not mean that there is automatically transferred, together with these functions, the abilities of independent thought, feeling, etc. Thought is a property of highly organized matter, the brain, its internal state, its capability to realize the

external world within itself, while distinguishing itself from the external world. This capability is not observed either in machines nor even in living things; it is the qualitatively distinguishing specific property, of the human being, this intelligence of his, and is observed only in embryo among the higher living organisms. If, in the human head, this property is accompanied by such actions as deduction, by some form of syllogism or another, and if these acts lend themselves to automation, then this certainly does not mean that the capability of the subject to think, to feel, to place before himself a definite goal, etc., lends itself in theory to automation as well. From the fact that some functions are transferred to the machine it is impossible to draw the hasty conclusion as to the possibility of transferring to it all the functions of the brain, including the ability of the brain to create and develop ways to transfer its functions to machines.

V.I. Lenin's "Materialism and Empiriocriticism," has been, and will continue to be, an inestimable aid to leaders, naturalists, philosophers, technologists and all our intelligentsia. In it, the modern reader finds the answers to those questions which were posed by the latest revolution in natural science 50 years ago and which today arise again and again in connection with the further progress of science and, also, in connection with the further stiffening of the philosophical struggle around the latest discoveries of physics, mathematics, automation and other domains of knowledge.

In his co-report, V. S. Pugachev spoke of the danger for specialists who use scientific concepts without analysing their philosophic side, particularly in those cases when these concepts are perverted in the spirit of Machism, which considers scientific laws only as convenient forms for registering experimental data. As an example of such Machian distortion, V. S. Pugachev considered certain definitions of basic concepts in probability theory, illustrating with them the actuality of V. I. Lenin's work in our time.

Among specialists there is widely disseminated the definition of probability of Mises as the limit of the ratio of the number of appearances of an outcome to the number of trials, as the number of trials increases without bound. This definition is completely unsound from the philosophical point of view and gives rise to great difficulties on the mathematical plane. Here, probability is not recognized as an objective characteristic of a random outcome, but as a convenient form for registering the result of performing test. In spite of this, probability theory allows many conclusions to be drawn without direct experiment.

The definitionof von Mises, despite its Machian character, imbues many applied works, for example, the book of the American scientists Laning and Battin "Random Processes in Automatic Control" (IL 1958), where it figures together with the correct logical foundations of probability theory.

Today, the fallaciousness of von Mises' definition of probability is known, among us, by almost all scientists. However, there are analogous Machian contents to the definitions of the correlation function and the spectral density of stationary random processes, widely disseminated both in foreign literature and in our own dealing with the technical applications of the theory of random processes.

In many sources today, the correlation function of a stationary random process, x(t), is defined by the formula

$$k(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\, x(t+\tau)\, dt. \tag{1}$$

This definition contains the same mathematical difficulty, in its passage to the limit, as does the definition of probability of von Mises. In this definition, the correlation function does not have a probabilistic content, is not an objective characteristic of the random process, but is essentially only a convenient form for registering experimental data. The correlation function does not exists only to the extent that one knows realizations, obtained experimentally, of the random process. Thus, the definition of the correlation function just given is, both from the mathematical and philosophical points of view, identically the von Mises definition of probability. Consequently, such a definition of the correlation function is purely Machian.

Whatever be the fallaciousness of the definition of the correlation function, formula (1) is correct in the particular case of an ergodic stationary random process, with the condition that the limit in it be understood as a limit in probability.

For the definition of the spectral density of a stationary random process, one ordinarily takes the Fourier transform of the "truncated" random process

$$A_T(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} x(t)\, e^{-i\omega t} dt, \tag{2}$$

and then, from the formula

$$\frac{1}{2T} \mid A_T(\omega) \mid^2 = \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\omega t} d\tau}{2T} \int_{-T}^{T} x(t)\, x(t+\tau)\, dt \tag{3}$$

the conclusion is drawn that, for all realizations of the ergodic stationary random function, the quantity

$$S(\omega) = \lim_{T\to\infty} \frac{1}{2T} \mid A_T(\omega) \mid^2 \tag{4}$$

has one and the same value, and formula (4) is then taken as the definition of the spectral density.

In fact, the passage to the limit in probability under the integral sign in (3) is impossible, and the quantity in (4) is essentially a random function, with different realizations for different realizations of the random function, with different realizations for different realizations of the random function x(t). Thus, formula (4) is invalid in either understanding of the limit. This example shows what harm can arise from the use of the Machian definitions and the refusal to penetrate deeply into the essence of the phenomena being studied. This error penetrates literally all the American books on the technical applications of random process theory, and many works of our scientists as well. In the book just cited of Laning and Battin, after a correct general definition of the spectral density of any random process, there is made the erroneous assertion that function (4) is one and the same for all realizations of the ergodic stationary random process.

Despite this, a somewhat deeper study of the question immediately leads to the opinion that function (4) must be different for different realizations of the random process x(t). Indeed, if we apply the inverse Fourier transformation to formula (2), we get

$$x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} A_\infty(\omega)\, e^{i\omega t} d\omega. \tag{5}$$

It is clear from this that the function $A_\infty(\omega)$ is random, and has different realization for different realizations of the random process x (t), since otherwise the function x (t) would be in no way random. This unconditionally provides the foundation for considering the function $S(\omega)$ random. A deeper investigation shows that the function $A_\infty(\omega)$ in (5) is white noise and, consequently, has an infinite dispersion. With this, function $S(\omega)$ has a spectral deviation of the same order as its mathematical expectation. Therefore, formula (4) should be corrected, and the spectral density of a stationary random process should be defined as the mathematical expectation of the function $S(\omega)$

$$S(\omega) = M[S(\omega)] = \lim_{T\to\infty} \frac{1}{2T} M[\mid A_T(\omega) \mid^2]. \tag{6}$$

What has been presented shows that it is important to develop scientific problems and concepts on the philosophical side, and the greatest aid in this matter is our book by V. I. Lenin.

A. A. Fel'dbaum, in his co-report, considered certain aspects of cybernetic machines from the philosophical

position. He analyzed individual links in the entire process, going from the human being's cognition of objective reality through the planning of controls to the direct control of the corresponding object.

The relationship of sensation and thought with the external world of objective reality, which exists independently of us, brilliantly investigated in the book by V. I. Lenin, has a direct relationship to cybernetics.

Cybernetics considers the processes occurring in the nervous system (sensations) and outside of us from a single point of view, showing that at their foundations are the material processes of transmission and transformation (processing) of information. This allows cybernetics to construct an analysis of the processes of sensation and thought on a material basis. Nevertheless, agnostics, and followers of Hume and Mach, deny in theory the possibility of studying the activities of the human brain by objective methods.

In rejecting the contention of the Machians that the brain is not the organ of thought, and thinking is not the function of the brain, Lenin remained at the position of Engels which is that thought and consciousness are the products of the human brain. Lenin says that the sophism of the idealist philosophers consists in this: that sensations are taken, not as connecting consciousness with the external world, but as partitions, walls, separating consciousness from the external world.

Thus, the materialistic approach discloses the relationships of sensation, consciousness with the phenomena outside of us. In the interpretation of cybernetics, this is expressed by the scheme of Fig. 1, where 1 and 2 are parts of the nervous system

The unity of the processes occurring in the human nervous system and in the world about us is contained in their materiality. The transmission and processing of information in different material systems, independently of their concrete material substrata, are subject to certain general laws, disclosed by cybernetics. The brain, the nervous system (the organs of thought) are not excluded from these relationships. Although the processes in the nervous system include other forms of motion than those in "dead" nature, the unity cited above and the connection of the links in the chain are valid, irrespective of their physical nature.

Mechanism does not raise itself above the recognition of the primitive unity of phenomena, and reduces the higher forms of motion to displacements. Dialectic materialism recognizes the qualitative diversity of motion, in considering motion as change in general. Between the qualitatively different forms of motion, despite their essential differences, there is a connection. This connection is expressed in cybernetics in the form of the unitary approach to the consideration of processes of signal processing, irrespective of the construction and physical nature of the links in the chain of stimuli.

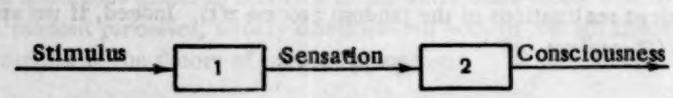


Fig. 1

The problem of cognition, which plays a large role in all sciences, is tremendously important in cybernetics as well. One can adduce an illustration from cybernetics of the Marxist-Leninist theory of cognition which consists of this: that the chain of stimuli (the chain of causal connections), in the terminology of cybernetics, is closed. From the point of view of cybernetics, the agnostic position may be presented as a rupturing of the closed chain connecting phenomena.

"Whence do we know that beyond "sensations" there stand actual (real) phenomena ?" says the agnostic The Marxist answers him that we can verify the truth of our representation in practice; by acting on nature, by observing and by carrying out experiments to predict, or even to give rise to phenomena. The mass of human activity convinces us that our representation mirrors an objective activity.

The human being knows by interacting with the world about him. The chain of cognition is closed via the activity of the human being (Fig. 2). Precisely in this closed circle does the process of cognition occur. The process of interaction in this circle is very complex and profound. The human being not only knows the world but also changes this world in the course of his activities, and is also changed himself.

"The human being would not be adapted to his environment if his sensations did not give him an objectively

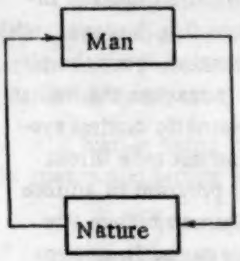


Fig. 2.

correct representation it" (Lenin's Collected Works, vol. 14, p. 166). A more detailed process of interaction of the human being with nature is considered in terms of the example of the activity of the investigator (Cf. Fig. 3). Let it be required to develop an optimal control system of some complex object. Initially, it is necessary to study it, to know its characteristics, at least to a first approximation. On the block diagram of Fig. 3, this is represented by the presence of the block, "Cognition of object."

For the cognition it is insufficient to observe the object passively, but it is generally necessary to effect actions on it, for example, to take down its characteristics. Such actions do not have the task of optimization of the object's regimen. These are the "cognitive" actions, experiments carried out on the object. The process of cognition does not reduce solely to the observation of reactions at the output to something at the input, since it is also necessary to construct hypotheses as to the structure of the object, and to verify them by acting on the object. This has greater value than the passive observation and collection of information, or even constructing tables of the relationships between inputs and outputs. The object can be considered as a "black box," as the thing in itself (der Ding im Sich), which, as a result of cognition, becomes "white" for us. When a definite degree of cognition is attained, the following step begins: the planning of strategies, of ways to exercise control. In the process of the activities of this step there may be needed more precise knowledge than before of the properties of the object (the line drawn from block "Planning" to block "Cognition of object" on Fig. 3). Then the control system is implemented, and the control block begins to actually act on the object. It may turn out that the results of this action are not qualitatively sufficient. Then, all this chain of actions will be repeated, but at an ever-higher level, until the required state of the control system is attained.

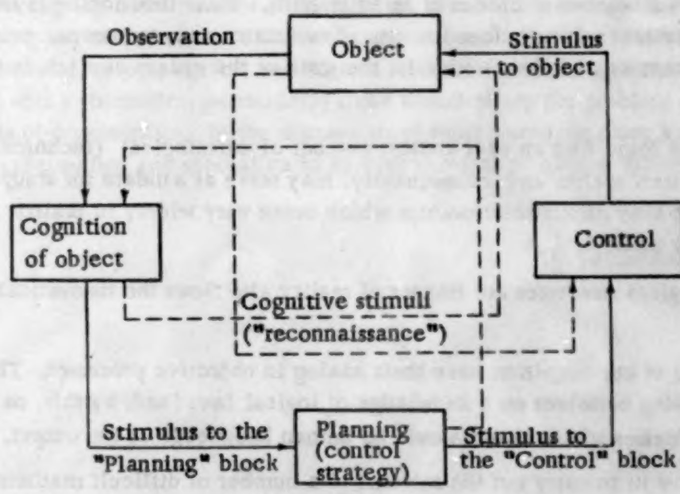


Fig. 3

The control system is called upon to replace the human being along certain segments of his intellectual activity. The structure of the complex automatic system of the future is given in the same form as the block schematic given above, but the operations in is are carried out without direct human intervention. In this system, the process of knowing (learning) the properties of the object will be simulated. This will be implemented by a "Cognition of object" block which, by observing the reactions of the object, by experimenting with it, by constructing hypotheses and checking them, will determine the probable characteristics of the object. Even today there exist "embryos" of this block: The reacting organ of an automatic control system. In a more developed form, this block is the self-adjusting system, carrying out trial motions – "experiments" – with the object and determining from their results the configuration of its characteristics. Such a system will contain both a block for "Planning of control" (in contemporary automata, its embryo is the control block) and a block for "Execution of control" (this is the executive organ, or even the whole automatic regulator, in modern automata).

As applied to the operation of automata, the word "cognition" should stand within quotation marks. Indeed, in the final analysis, it is not the automation that knows, but the human being: he does this, however, with the collaboration of his tools such as machines, including machines for control. The information about the object's characteristics collected by the automaton only enriches the cognition process when it reaches the human being. The fact of the matter is that, if the scheme of Fig. 3 were the schematic of an automatic control system, it would be necessary to add lines to it connecting it to a human being. The latter can not take direct part in the control processes of this system, but it was he who constructed it, who placed its program of actions in it, i.e., he acted on it. Further, the results of the operation of this system, destined for human beings, are supplied to them in acting on them. Thus, the total structure of the interconnections of an automatic system ineluctably includes the human being as the central link.

S. M. Shalyutin, in his co-report, showed that the idea of mechanizing intellectual tasks, and using machines for solving logical problems, is profoundly materialistic. Not only the content, but also the form, of thought are mirrored in the external world.

The relationships of concepts, the determination of the correct forms of these relationships, this being one of the important conditions for true thought, is a reflection of general and stable relationships between things. Lenin showed that the "logical figures" are the most common relationships between things. These relationships of things are imprinted on the consciousness of human beings in the guise of forms of thought in the processes of practical human activity. Therefore, the study of the forms of thought may, and must have a definite value even outside of logic.

The basic laws of formal logic have as their principal ontological content the relative stability of things, their qualitative determinateness, etc.

A developed logical apparatus may serve as a powerful instrument in the investigation and synthesis of those objective systems in whose functioning the ontological content of logical forms and structures plays an essential role, in particular, a number of classes of relay systems. Their functioning is subject to the law of the excluded middle: each relay is of necessity found in one of two states and, moreover, only in one of them. The same is also true of the system as a whole. With this, the state of the system as a whole is a function of the states of its elements.

The abstract results of logic find an ever greater number of extralogical (technical and other) applications, first, because they mirror reality and, consequently, may serve as a means for studying that reality indirectly; secondly, because they mirror relationships which occur very widely in reality, and may therefore he interpreted in multitudinous ways.

From the fact that logical structures are images of reality also flows the theoretical possibility of the existence of logical machines.

The content and form of our cognition have their analog in objective processes. There is therefore the objective possibility, by basing ourselves on a knowledge of logical laws (and, by this, on certain laws of reality), of artifically organizing processes whose results would be human knowledge at the output.

Today, machines allow us to carry out the solution of a number of difficult mathematical problems, problems of controlling production. From these achievements, cybernetics draws certain false conclusions, in particular, thought is identified with processes in automata. Indeed, there is a definite likeness between automata and thinking beings, but this likeness cannot be rendered absolute; it is not an identity.

The problem of distinguishing between concept formation in human consciousness, and the symbolization of constant properties of an external field by an automation, is a complex problem containing a number of aspects. Consciousness is a social product. In the process of working, the human being changes nature in his interests, and from this stems the purposeful activity, without which there would be no work processes. To attain his goals, the human being uses the materials of labor. And any cybernetic system remains in the productive process in the function of a means of labor, because production does not alter a natural thing in its interests, and it is not from it that, in the final analysis, the purpose of the process stems. This system is only a means for reaching a goal.

People in the processes of production enter into economic relationships with one another.

On this basis, the ideology is formulated. The human consciousness not only fixes the presence of definite orders, but also includes in itself the relationships to them. An automaton, even though capable of symbolizing certain sides of societal activity, cannot "relate" to them. It is meaningless to speak of the opinions, the ideologies, of an automaton.

A human being in the process of his practical activity becomes a subject, an "I", separating himself from the means and setting himself up in opposition to them.

The objective opposition of man and nature, which is realized in labor, also gives birth to the opposition of nature and man in the guise of subject and object. An automaton is not a subject, but an objective material element of production, in opposition to the human subject. Thought is a function of the human brain — the product of a long biological and societal history.

The aspects just considered may serve as approaches to the problem of distinguishing thought from the processes in an automaton. Insofar as the machine does not think, it is meaningless to argue as to whether or not a machine might be cleverer than man, its creator.

Man, equipped with machines, is capable of solving problems which he would be incapable of solving without them. From the fact that machines cannot think, it does not at all follow that the machine can solve only very simple problems. The capabilities of cybernetic systems, constructed by technicians, are impossible to determine simply by starting from the fact that the machine is not a subject.

From this point of view, the attempts to limit cybernetics are unfounded. From the fact that technological cybernetic systems do not think, it does not follow at all that they might not be used for simulating definite sides of thought processes. A similar aspect is presented by the one-sided absolute difference imputed to exist between thinking beings and automata. It is not less harmful than the imputation of an absolute likeness between them.

In the process of developing new branches of knowledge, cybernetics among them, there is great value in the methodological working out of questions. The Academic Council adopted a resolution to continue work in this direction and, in particular, to prepare for, and hold, a scientific colloquium for discussing the most important questions connected with cybernetics, particularly those which affect the problem of constructing the material-technological basis of communism. In the discussions of these questions there will take part, both specialists in the domain of cybernetics, and specialists in various domains of natural science and philosophy.

D. Ya. Libenson